GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: July 25, 2001, 04:54:54 ; Search time 2762.24 Seconds
(without alignments)
65.021 Million cell updates/sec

Title: US-09-142-095-4
Perfect score: 19
Sequence: 1 tttgctcctgccagaggtt 19

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est33:*
26: gb_est34:*
27: gb_est35:*
28: gb_est36:*
29: gb_est37:*
30: gb_est38:*
31: gb_est39:*
32: gb_est40:*
33: em_estba:*
34: em_estfun:*
35: em_esthum1:*
36: em_esthum2:*
37: em_esthum3:*
38: em_esthum4:*
39: em_esthum5:*
40: em_esthum6:*
41: em_esthum7:*
42: em_esthum8:*
43: em_esthum9:*
44: em_esthum10:*
45: em_esthum11:*
46: em_esthum12:*
47: em_esthum13:*
48: em_esthum14:*
49: em_esthum15:*
50: em_esthum16:*
51: em_esthum17:*
52: em_esthum18:*
53: em_esthum19:*
54: em_esthum20:*
55: em_esthum21:*
56: em_esthum22:*
57: em_esthum23:*
58: em_esthum24:*
59: em_esthum25:*
60: em_esthum26:*
61: em_esthum27:*
62: em_esthum28:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estin5:*
68: em_estom1:*
69: em_estom2:*
70: em_estov1:*
71: em_estov2:*
72: em_estpl1:*
73: em_estpl2:*
74: em_estpl3:*
75: em_estpl4:*
76: em_estpl5:*
77: em_estpl6:*
78: em_estpl7:*
79: em_estpl8:*
80: em_estpl9:*
81: em_estpl10:*
82: em_estro1:*
83: em_estro2:*
84: em_estro3:*
85: em_estro4:*
86: em_estro5:*
87: em_estro6:*
88: em_estro7:*
89: em_estro8:*
90: em_estro9:*
91: em_estro10:*
92: em_estro11:*
93: em_estro12:*
94: em_estro13:*
95: em_estro14:*
96: em_estro15:*
97: em_estro16:*
98: em_estro17:*
99: em_estro18:*
100: em_estro19:*
101: em_estro20:*
102: gb_est25:*
103: gb_est26:*
104: gb_est27:*
105: gb_est28:*
106: gb_est29:*
107: gb_est30:*
108: gb_est31:*
109: gb_est32:*
110: gb_est41:*
111: gb_est42:*
112: gb_est43:*
113: gb_est44:*
114: gb_est45:*
115: gb_est46:*
116: gb_est47:*

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est75:*
137: gb_est76:*
138: gb_est77:*
139: gb_est78:*
140: gb_est79:*
141: gb_est80:*
142: gb_est81:*
143: gb_est82:*
144: gb_est83:*
145: gb_est84:*
146: gb_est85:*
147: gb_est86:*
148: gb_est87:*
149: gb_est88:*
150: gb_est89:*
151: gb_est90:*
152: gb_est99:*
153: gb_est100:*
154: gb_est101:*
155: gb_est102:*
156: gb_est103:*
157: gb_est104:*
158: gb_est105:*
159: gb_est106:*
160: gb_est67:*
161: gb_est68:*
162: gb_est69:*
163: gb_est70:*
164: gb_est71:*
165: gb_est72:*
166: gb_est73:*
167: gb_est74:*
168: gb_est91:*
169: gb_est92:*
170: gb_est93:*
171: gb_est94:*
172: gb_est95:*
173: gb_est96:*
174: gb_est97:*
175: gb_est98:*
176: em_esthum29:*
177: em_esthum30:*
178: em_esthum31:*
179: em_esthum32:*
180: em_esthum33:*
181: em_estom3:*
182: em_estpl11:*
183: em_estro21:*
184: em_estro22:*
185: em_estro23:*
186: em_htc:*
187: gb_est107:*
188: gb_est108:*
189: gb_est109:*
190: gb_est110:*
191: gb_est111:*
192: gb_htc:*
193: em_gss_fun:*
194: em_gss_hum1:*
195: em_gss_hum2:*
196: em_gss_hum3:*
197: em_gss_hum4:*
198: em_gss_hum5:*
199: em_gss_hum6:*
200: em_gss_hum7:*
201: em_gss_hum8:*
202: em_gss_hum9:*
203: em_gss_inv1:*
204: em_gss_inv2:*
205: em_gss_inv3:*
206: em_gss_other:*
207: em_gss_pln1:*
208: em_gss_pln2:*
209: em_gss_pro:*
210: em_gss_rod1:*
211: em_gss_rod2:*
212: em_gss_rod3:*
213: em_gss_rod4:*
214: em_gss_rod5:*
215: em_gss_vrt1:*
216: em_gss_vrt2:*
217: em_gss_vrt3:*
218: gb_gss1:*
219: gb_gss2:*
220: gb_gss3:*
221: gb_gss4:*
222: gb_gss5:*
223: gb_gss6:*
224: gb_gss7:*
225: gb_gss8:*
226: gb_gss9:*
227: gb_gss10:*
228: gb_gss11:*
229: gb_gss12:*
230: gb_gss13:*
231: gb_gss14:*
232: gb_gss15:*
233: gb_gss16:*
234: gb_gss17:*
235: gb_gss18:*
236: gb_gss19:*
237: gb_gss20:*
238: gb_gss21:*
239: gb_gss22:*
240: gb_gss23:*
241: gb_gss24:*
242: gb_gss25:*
243: gb_gss26:*
244: gb_gss27:*
245: gb_gss28:*
246: gb_gss29:*
247: gb_gss30:*
248: gb_gss31:*
249: gb_gss32:*
250: gb_gss33:*
251: gb_gss34:*
252: em_gss_inv4:*
253: em_gss_rod6:*
254: em_gss_rod7:*
255: em_gss_rod8:*
256: gb_gss35:*
257: gb_gss36:*
258: gb_gss37:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	17.4	91.6	522	150	BF545097	BF545097 UI-R-C2p-
c 2	17.4	91.6	608	230	AQ540572	AQ540572 RPCI-11-3
3	17	89.5	367	115	AW429811	AW429811 68306 MAR
c 4	17	89.5	936	175	BG297415	BG297415 602395559
5	16.4	86.3	356	223	AQ068059	AQ068059 HS_2200_B
6	16.4	86.3	530	233	AQ799987	AQ799987 HS_5300_B
c 7	16.4	86.3	574	104	AI979827	AI979827 pat.pk000
8	16.4	86.3	616	247	AZ664475	AZ664475 1M0544H08
9	16.4	86.3	680	141	BE901069	BE901069 601674415
c 10	16.4	86.3	978	173	BG109825	BG109825 602280983
11	16	84.2	206	17	AI203923	AI203923 qd72h07.x
12	16	84.2	338	8	AA504455	AA504455 aa60a05.s
13	16	84.2	341	9	AA577198	AA577198 nm87b11.s
14	16	84.2	738	233	AQ751483	AQ751483 HS_5576_B
c 15	15.8	83.2	115	190	W25430	W25430 zb90c01.r1
c 16	15.8	83.2	132	19	AI361553	AI361553 qy44b11.x
c 17	15.8	83.2	133	19	AI363101	AI363101 qy54h09.x
c 18	15.8	83.2	165	163	BE142610	BE142610 IL0-HT015
c 19	15.8	83.2	192	114	AW306100	AW306100 fj79h04.y
c 20	15.8	83.2	243	161	BB574964	BB574964 BB574964
c 21	15.8	83.2	251	161	BB580263	BB580263 BB580263
c 22	15.8	83.2	264	112	AW159304	AW159304 za63c09.x
23	15.8	83.2	287	116	AW482661	AW482661 46483 MAR
c 24	15.8	83.2	287	128	BB218963	BB218963 BB218963
25	15.8	83.2	290	134	BB466184	BB466184 BB466184
c 26	15.8	83.2	332	118	AW583615	AW583615 ia02a02.y
c 27	15.8	83.2	362	169	BF802628	BF802628 CM0-CI013
c 28	15.8	83.2	388	6	AA378262	AA378262 EST90949
c 29	15.8	83.2	388	23	AI639055	AI639055 rx04126s
30	15.8	83.2	401	170	BF857435	BF857435 QV3-FT021
c 31	15.8	83.2	417	5	AA309827	AA309827 EST180840
c 32	15.8	83.2	428	190	W38798	W38798 zb27e08.r1
c 33	15.8	83.2	430	175	BG234717	BG234717 dab63b10.
34	15.8	83.2	455	244	AZ477763	AZ477763 1M0297J23
35	15.8	83.2	464	228	AQ436361	AQ436361 HS_5060_B
c 36	15.8	83.2	481	1	AA036818	AA036818 zk29a07.r
c 37	15.8	83.2	487	159	N42389	N42389 yy09a02.r1
38	15.8	83.2	494	169	BF743356	BF743356 IL2-BT080
c 39	15.8	83.2	495	24	AI722856	AI722856 fc32e10.y
c 40	15.8	83.2	502	229	AQ484899	AQ484899 RPCI-11-2
c 41	15.8	83.2	504	105	AL049145	AL049145 DKFZp434F
c 42	15.8	83.2	515	111	AW078344	AW078344 fe48c07.y
c 43	15.8	83.2	521	233	AQ772645	AQ772645 HS_5566_A
c 44	15.8	83.2	525	235	AQ917005	AQ917005 nbeb0067L
45	15.8	83.2	549	230	AQ533974	AQ533974 RPCI-11-3

ALIGNMENTS

RESULT 1
BF545097
LOCUS BF545097 522 bp mRNA EST 11-DEC-2000
DEFINITION UI-R-C2p-qt-e-07-0-UI.r1 UI-R-C2p Rattus norvegicus cDNA clone UI-R-C2p-qt-e-07-0-UI 5', mRNA sequence.
ACCESSION BF545097
VERSION BF545097.1 GI:11636204
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 522)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
cDNA Library Preparation: M.B. Soares Lab Clone distribution: clones will be available through Research Genetics (www.resgen.com) This clone is also available through the I.M.A.G.E. Consortium at LLNL (info@image.llnl.gov). IMAGE ID= 1789283
Seq primer: M13 Forward.
FEATURES Location/Qualifiers
source 1. .522
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-C2p-qt-e-07-0-UI"
/clone_lib="UI-R-C2p"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-C2p library is a subtracted library derived from the UI-R-C1 library, which is a subtracted library derived from the UI-R-C0 library. The UI-R-C0 library consisted of a mixture of individually tagged normalized libraries constructed from rat placenta, adult lung, brain, liver, kidney, heart, spleen, ovary, muscle, 8, 12 and 18-day embryo. The tag is a string of 3-5 nucleotides present between the Not I site and the oligo-dT track which allows identification of the library of origin of a clone within the mixture. The subtracted library (UI-R-C2p) was constructed as follows: PCR amplified cDNA inserts from UI-R-C1 clones from which 3' ESTs had been derived was used as a driver in a hybridization with the UI-R-C1 library in the form of single-stranded circles. The remaining single-stranded circles (subtracted library) was purified by hydroxyapatite column chromatography, converted to double-stranded circles and electroporated into DH10B bacteria (Life Technologies) to generate the UI-R-C2p library. This procedure has been previously described (Bonaldo, Lennon and Soares, Genome Research 6: 791-806, 1996)"
BASE COUNT 156 a 117 c 104 g 145 t
ORIGIN

Query Match 91.6%; Score 17.4; DB 150; Length 522;
Best Local Similarity 94.7%; Pred. No. 2.1e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy      1 tttgctcctgccagaggtt 19
          || |||||||||||||||||
Db    128 TTGGCTCCTGCCAGAGGTT 146
```

RESULT 2
AQ540572/c
LOCUS AQ540572 608 bp DNA GSS 19-MAY-1999
DEFINITION RPCI-11-358E23.TV RPCI-11 Homo sapiens genomic clone RPCI-11-358E23, DNA sequence.
ACCESSION AQ540572
VERSION AQ540572.1 GI:4871102
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 608)
AUTHORS Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter

```
                ,J.C.
  TITLE         Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
                Map Building
  JOURNAL       Unpublished (1997)
COMMENT         Contact: Shaying Zhao, William Nierman, Mark Adams
                Department of Eukaryotic Genomics
                The Institute for Genomic Research
                9712 Medical Center Dr., Rockville, MD 20850
                Tel: 301 838 0200
                Fax: 301 838 0208
                Email: hbe@tigr.org
                Clones are derived from the human BAC library RPCI-11. For BAC
                library availability, please contact Pieter de Jong
                (pieter@dejong.med.buffalo.edu). Clones may be purchased from
                BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
                Research Genet cs (info@resgen.com). BAC end search page:
                http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
                Seq primer: T7
                Class: BAC ends.
FEATURES                Location/Qualifiers
     source             1. .608
                        /organism="Homo sapiens"
                        /db_xref="GDB:7637206"
                        /db_xref="taxon:9606"
                        /clone="RPCI-11-358E23"
                        /clone_lib="RPCI-11"
                        /sex="Male"
                        /cell_type="Lymphocytes"
                        /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                        RPCI11 Human Male BAC Library"
BASE COUNT       259 a     83 c     93 g    173 t
ORIGIN


  Query Match             91.6%;  Score 17.4;  DB 230;  Length 608;
  Best Local Similarity   94.7%;  Pred. No. 2.1e+02;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 tttgctcctgccagaggtt 19
           || ||||||||||||||||
Db     138 TTGGCTCCTGCCAGAGGTT 120


RESULT   3
AW429811
LOCUS       AW429811      367 bp    mRNA            EST       09-JUL-2000
DEFINITION  68306 MARC 1PIG Sus scrofa cDNA 5', mRNA sequence.
ACCESSION   AW429811
VERSION     AW429811.1  GI:6961118
KEYWORDS    EST.
SOURCE      pig.
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 367)
  AUTHORS   Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,
            Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W.
            and Keele,J.W.
  TITLE     Design and use of two pooled tissue normalized cDNA libraries for
            EST discovery in swine
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 20
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 28  row: C  column: 15
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES                Location/Qualifiers
     source             1. .367
                        /organism="Sus scrofa"
                        /db_xref="taxon:9823"
                        /clone_lib="MARC 1PIG"
                        /tissue_type="pooled"
                        /lab_host="DH10B"
                        /note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
                        Library made from pooled tissue from day 11, 13, 15, 20,
                        and 30 embryos."
BASE COUNT       91 a    100 c    115 g     60 t      1 others
ORIGIN


  Query Match             89.5%;  Score 17;  DB 115;  Length 367;
  Best Local Similarity   100.0%;  Pred. No. 3.1e+02;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 tgctcctgccagaggtt 19
           |||||||||||||||||
Db     256 TGCTCCTGCCAGAGGTT 272


RESULT   4
BG297415/c
LOCUS       BG297415      936 bp    mRNA            EST       21-FEB-2001
DEFINITION  602395559F1 NIH_MGC_94 Mus musculus cDNA clone IMAGE:4507022 5',
            mRNA sequence.
ACCESSION   BG297415
VERSION     BG297415.1  GI:13061044
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 936)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: The Cepko Laboratory
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10383  row: f  column: 15
            High quality sequence stop: 604.
FEATURES                Location/Qualifiers
     source             1. .936
                        /organism="Mus musculus"
                        /db_xref="taxon:10090"
                        /clone="IMAGE:4507022"
                        /clone_lib="NIH_MGC_94"
                        /tissue_type="retina"
                        /lab_host="DH10B (phage-resistant)"
                        /note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
                        Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                        Average insert size 3.3 kb. Library enriched for
                        full-length clones and constructed by Life Technologies.
                        Note: this is a NIH_MGC Library."
BASE COUNT      367 a    218 c    259 g     92 t
ORIGIN


  Query Match             89.5%;  Score 17;  DB 175;  Length 936;
  Best Local Similarity   100.0%;  Pred. No. 3.6e+02;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: July 25, 2001, 05:23:03 ; Search time 247.87 Seconds
(without alignments)
53.197 Million cell updates/sec

Title: US-09-142-095-2
Perfect score: 21
Sequence: 1 ccactgggatcaacagtatct 21

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_0601:*
1: /SIDS1/gcgdata/geneseq/geneseqn/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	21	100.0	21	18	AAT79542	UGT1*1 gene exon 1
c	2	21	100.0	864	21	AAZ45110	UDP-glucuronosyltr
c	3	21	100.0	1167	13	AAQ33024	UGT1A Exon 1 from
c	4	21	100.0	2351	13	AAQ27369	HUG-Br1. Homo sap
	5	15.8	75.2	504	21	AAD00564	Streptococcus pyog
	6	15.8	75.2	697	20	AAX30352	DNA encoding a hum
	7	15.8	75.2	1073	19	AAV30310	Bacillus thuringie
	8	15.8	75.2	4351	19	AAV70397	LRP5 isoform 2 lon
	9	15.8	75.2	4843	19	AAV70395	LRP5 cDNA longest
	10	15.8	75.2	4915	19	AAV70398	LRP5 isoform 3 put
	11	15.8	75.2	5022	19	AAV85549	LRP5 isoform 5 cDN
	12	15.8	75.2	5098	19	AAV70396	LRP5 isoform 1 cDN
	13	15.8	75.2	5125	19	AAV86019	Lrp5 isoform 3 cDN
	14	15.8	75.2	5162	19	AAV85550	LRP5 isoform 6 cDN
	15	15.8	75.2	5166	19	AAV85551	LRP5 isoform 2 cDN
	16	15.8	75.2	5263	19	AAV70400	LRP5 isoform 4 cDN
c	17	15.4	73.3	184	13	AAQ27833	Sequence which hyb
	18	15.4	73.3	1208	21	AAA87688	Human secreted pro
	19	15.2	72.4	82	12	AAQ14774	Casein signal pept
	20	15.2	72.4	274	21	AAA43269	Xenopus secreted e
c	21	15.2	72.4	652	20	AAV88862	EST clone HZ15. H
	22	15.2	72.4	691	20	AAX39707	Gastric cancer ass
	23	15.2	72.4	1210	12	AAQ14051	Rennin gene. Synt
	24	15.2	72.4	1312	21	AAC77235	Human ORFX ORF2790
	25	15.2	72.4	1416	20	AAZ17647	Human gene express
c	26	15.2	72.4	1900	10	AAN92624	Nucleotide sequenc
c	27	15.2	72.4	7498	21	AAA63786	Nucleotide sequenc
	28	15.2	72.4	8930	19	AAV22834	Haemophilus paraga
	29	15.2	72.4	20682	20	AAX20569	Polynucleotide seq
c	30	15	71.4	434	21	AAC05486	Human secreted pro
c	31	15	71.4	715	20	AAV69488	Banana fruit ripen
	32	14.8	70.5	260	21	AAA44168	Human secreted exp
	33	14.8	70.5	297	21	AAA15919	Human protein clon
c	34	14.8	70.5	300	20	AAZ13302	Human gene express
	35	14.8	70.5	354	20	AAX41509	Human secreted pro
	36	14.8	70.5	379	21	AAA41312	Human secreted exp
	37	14.8	70.5	392	21	AAA16090	Human colon cancer
	38	14.8	70.5	413	20	AAX97873	Human secreted pro
	39	14.8	70.5	415	21	AAC00147	Human secreted pro
	40	14.8	70.5	459	21	AAA38007	UCK-2 nucleotide s
	41	14.8	70.5	467	21	AAA15929	Human protein clon
	42	14.8	70.5	495	20	AAZ34052	Human EST DNA43509
	43	14.8	70.5	495	21	AAC78499	Human EST DNA43509
	44	14.8	70.5	495	22	AAF93358	cDNA encoding SRT
	45	14.8	70.5	500	20	AAX97826	Human secreted pro

ALIGNMENTS

RESULT 1
AAT79542
ID AAT79542 standard; DNA; 21 BP.
XX
AC AAT79542;
XX
DT 23-JAN-1998 (first entry)
XX
DE UGT1*1 gene exon 1 upstream PCR primer B.
XX
KW Uridine diphosphate glucuronosyltransferase gene; UGT;
KW Gilbert's syndrome; GS; unconjugated hyperbilirubinaemia;
KW bilirubin glucuronidation; Crigler-Najjar; type 2; drug metabolism;
KW Drug trial efficiency; screening; PCR primer; ss.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN WO9732042-A2.
XX
PD 04-SEP-1997.
XX
PF 03-MAR-1997; 97WO-GB00577.
XX
PR 16-MAR-1996; 96GB-0005598.
PR 01-MAR-1996; 96GB-0004480.
XX
PA (UYDU-) UNIV DUNDEE.
XX
PI Burchell B;
XX
DR WPI; 1997-448702/41.
XX
PT Improving drug trial efficiency comprises identifying participants

PT with Gilbert's syndrome - useful as their altered drug metabolism
PT may hinder result interpretation
XX
PS Claim 14; Page 12; 31pp; English.
XX
CC This PCR primer (with primers AAT79541-44) flanks the TATA box sequence
CC upstream of the uridine diphosphate glucuronosyltransferase (UGT) gene
CC 1*1 exon 1 (see AAT79540), and was used to amplify fragments of 253-255
CC bp. This gene is known to be associated with Gilbert's syndrome (GS). GS
CC is a mild, common form of unconjugated hyperbilirubinaemia associated
CC with reduced bilirubin glucuronidation capacity. Analysis of the genetic
CC basis of GS has allowed 2 forms to be identified. One is a mild form
CC associated with a homozygous 2 bp insertion in the TATA sequence
CC upstream of the UGT1*1 exon 1, and the other is a more severe form
CC associated with heterozygosity for a mutation which, when homozygous,
CC causes Crigler-Najjar type 2 disease. The first form is autosomal
CC recessive and the second is inherited dominantly. Patients suffering from
CC GS, which is benign, may have altered metabolism of some drugs, making it
CC difficult to determine if an effect is due to the drug or the syndrome.
CC Drug trial efficiency would be improved if potential participants can be
CC screened for the genetic basis of GS, and eliminated or included on
CC basis of them possessing or not possessing GS.
XX
SQ Sequence 21 BP; 6 A; 6 C; 4 G; 5 T; 0 other;

Query Match 100.0%; Score 21; DB 18; Length 21;
Best Local Similarity 100.0%; Pred. No. 0.11;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ccactgggatcaacagtatct 21
||||||||||||||||||||
Db 1 ccactgggatcaacagtatct 21

RESULT 2
AAZ45110/c
ID AAZ45110 standard; DNA; 864 BP.
XX
AC AAZ45110;
XX
DT 28-FEB-2000 (first entry)
XX
DE UDP-glucuronosyltransferase 1 (UGT1) exon 1A nucleotide sequence.
XX
KW Uridine diphosphate-glucuronosyltransferase 1; UGT1; polymorphism; probe;
KW glucuronic acid; Crigler-Najjar syndrome; Gilbert syndrome; jaundice;
KW unconjugated hyperbilirubinaemia; drug metabolism; transgenic animal;
KW pharmacogenetic screening; diagnose; ss.
XX
OS Homo sapiens.
XX
PN WO9957322-A2.
XX
PD 11-NOV-1999.
XX
PF 04-MAY-1999; 99WO-US09702.
XX
PR 07-MAY-1998; 98US-0084807.
XX
PA (AXYS-) AXYS PHARM INC.
XX
PI Penny L, Galvin M;
XX
DR WPI; 2000-052981/04.
DR P-PSDB; AAY57093.
XX
PT New nucleic acid representing polymorphisms in the human uridine
PT diphosphate glucuronosyltransferase gene, used for diagnosis and
PT evaluation of drug metabolism -
XX
PS Examples; Page 27-29; 63pp; English.
XX
CC Nucleotide sequences AAZ45110-Z45118 are exons 1A-1J of the human
CC uridine diphosphate-glucuronosyltransferase 1 (UGT1) gene. The UGTs are a
CC family of enzymes that catalyse the glucuronic acid conjugation of a wide
CC range of endogenous and exogenous substrates including phenols,
CC alcohols, amines and fatty acids. Many of the reactions catalysed by UGTs
CC result in toxic substances being converted to compounds which are more
CC water soluble and are excreted. The invention relates to and identifies
CC UGT1 polymorphisms (AAZ45004-Z45041). The polymorphism sequences are
CC useful as probes for detecting UGT1 locus polymorphisms, indicative of
CC altered UGT1 expression or activity. These polymorphisms are associated
CC with Crigler-Najjar and Gilbert syndromes (unconjugated
CC hyperbilirubinaemia) and drug metabolism. The genotyping of the UGT1 gene
CC is used to predict the rate of metabolism of UGT1 substrates, possible
CC drug-drug interactions and adverse side effects (i.e. to optimize drug
CC dosage), and to screen for diseases caused by exposure to toxins and to
CC study the effects of polymorphisms on enzymatic activity. The UGT1
CC sequences, including polymorphisms, can also be used to produce the
CC corresponding protein (or its fragments) or to generate transgenic
CC animals or modified cells e.g. for pharmacogenetic screening.
XX
SQ Sequence 864 BP; 185 A; 236 C; 209 G; 234 T; 0 other;

Query Match 100.0%; Score 21; DB 21; Length 864;
Best Local Similarity 100.0%; Pred. No. 0.18;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ccactgggatcaacagtatct 21
||||||||||||||||||||
Db 106 CCACTGGGATCAACAGTATCT 86

RESULT 3
AAQ33024/c
ID AAQ33024 standard; DNA; 1167 BP.
XX
AC AAQ33024;
XX
DT 27-JAN-1993 (first entry)
XX
DE UGT1A Exon 1 from the UGT1 gene locus.
XX
KW UGT1A; UGT1BP; UGT1C; UGT1D; UGT1E; UGT1F; isozyme; bilirubin;
KW UDP-glucuronosyl transferase; CN; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT misc_RNA 1..20
FT /*tag= a
FT /note= "representation of 11.7 kbp of
FT non-sequenced DNA between the sequences
FT represented in AAQ33023 and AAQ33024"
FT TATA_signal 52..66
FT /*tag= b
FT exon 166..909
FT /*tag= c
FT /label= UGT1A_Exon_1
FT misc_RNA 181..909
FT /*tag= d
FT /note= "encodes transferase isoform; see CC"
FT misc_RNA 1148..1167
FT /*tag= e
FT /note= "representation of 5 kbp of
FT non-sequenced DNA between the sequences
FT represented in AAQ33024 and AAQ33025"
XX
PN WO9212987-A.
XX
PD 06-AUG-1992.
XX
PF 10-JAN-1992; 92WO-US00282.
XX

```
PR   10-JAN-1991;   91US-0639453.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICE.
XX
PI   Owens IS,  Ritter JK;
XX
DR   WPI; 1992-284593/34.
DR   P-PSDB; AAR30194.
XX
PT   Isolated gene locus UGT1, DNA segments and diagnostic probes -
PT   for diagnosing Gilbert's disease and Crigler-Najjar syndrome
PT   types I and II
XX
PS   Disclosure; Fig 1F; 99pp; English.
XX
CC   The isolated gene locus, UGT1, has a sequence of about 10000 bp
CC   which represent (1) Exon 1, comprising 6 transcriptional units
CC   (UGT1F, E, D, C, BP and A), represented in AAQ27368 and
CC   AAQ33020-24 respectively;
CC                  (2) Exon 2, represented in AAQ33025;
CC                  (3) Exon 3, represented in AAQ33026;
CC                  (4) Exon 4, represented in AAQ33026;
CC                  (5) Exon 5, represented in AAQ33027; and
CC                  (6) about 69 kb of non-sequenced DNA.
CC   Six unique N-termini of 286-289 amino acids are encoded by
CC   the six different first exons and identical C-termini of 246 amino
CC   acids are encoded by the common exons 2-5.  The UGT1 gene locus
CC   encodes a family of UDP-glucuronosyl transferase isozymes, two of
CC   which metabolise bilirubin.
CC   Patients having Crigler-Najjar Syndrome (CN) Type I, have a
CC   mutation present in the second common exon.
XX
SQ   Sequence 1167 BP; 255 A; 259 C; 272 G; 340 T; 41 other;


  Query Match           100.0%;  Score 21;  DB 13;  Length 1167;
  Best Local Similarity   100.0%;  Pred. No. 0.19;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ccactgggatcaacagtatct 21
          |||||||||||||||||||||
Db    286 CCACTGGGATCAACAGTATCT 266


RESULT   4
AAQ27369/c
ID   AAQ27369 standard; cDNA; 2351 BP.
XX
AC   AAQ27369;
XX
DT   27-JAN-1993  (first entry)
XX
DE   HUG-Br1.
XX
KW   Bilirubin; UDP-glucuronosyltransferase; HUGBr1; HUGBr2;
KW   monoglucuronide; diglucuronide; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             16..784
FT                   /*tag=  a
FT   polyA_signal    2330..2335
FT                   /*tag=  b
FT                   /number= 1
FT   polyA_signal    2338..2343
FT                   /*tag=  c
FT                   /number= 2
XX
PN   WO9212987-A.
XX
PD   06-AUG-1992.
XX
PF   10-JAN-1992;   92WO-US00282.
XX
PR   10-JAN-1991;   91US-0639453.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICE.
XX
PI   Owens IS,  Ritter JK;
XX
DR   WPI; 1992-284593/34.
DR   P-PSDB; AAR26153.
XX
PT   Isolated gene locus UGT1, DNA segments and diagnostic probes -
PT   for diagnosing Gilbert's disease and Crigler-Najjar syndrome
PT   types I and II
XX
PS   Disclosure; Fig 9A-I; 99pp; English.
XX
CC   Two human liver bilirubin UDP-glucuronosyltransferase cDNAs have
CC   been isolated.  They are referred to as HUGBr1 (AAQ27369) and HUGBr2
CC   (AAQ27370) (Ritter, et al., J. Biol. Chem. 266:1043-1047 (1991)) and,
CC   upon expression individually in COS-1 cells, encode isoforms that
CC   catalyse the formation of the two bilirubin monoglucuronides and
CC   the diglucuronide.
CC   The cDNAs contain identical 3' ends (1469 bp in length) to each
CC   other and to that of the human phenol transferase cDNA, HLUGP1
CC   (Harding et al., Proc. Natl. Aca. Sci. USA 85:8281 (1988)).
CC   In contrast, they have unique 5' ends.
XX
SQ   Sequence 2351 BP; 602 A; 540 C; 556 G; 653 T; 0 other;


  Query Match           100.0%;  Score 21;  DB 13;  Length 2351;
  Best Local Similarity   100.0%;  Pred. No. 0.21;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ccactgggatcaacagtatct 21
          |||||||||||||||||||||
Db    121 CCACTGGGATCAACAGTATCT 101


RESULT   5
AAD00564
ID   AAD00564 standard; DNA; 504 BP.
XX
AC   AAD00564;
XX
DT   29-AUG-2000  (first entry)
XX
DE   Streptococcus pyogenes strain KTL3 partial GRAB protein encoding DNA.
XX
KW   GRAB protein; protein G related alpha2M binding protein; vaccine;
KW   alpha2-macroglobulin; group A Streptococcus; GAS; antibiotic;
KW   immune response; Streptococcus pyogenes infection; ds.
XX
OS   Streptococcus pyogenes.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..504
FT                   /*tag=  a
FT                   /product= "GRAB protein"
FT                   /partial
XX
PN   WO200026240-A2.
XX
PD   11-MAY-2000.
XX
PF   02-NOV-1999;   99WO-GB03631.
XX
PR   02-NOV-1998;   98GB-0023975.
XX
PA   (ACTI-) ACTINOVA LTD.
XX
PI   Bjorck LH,  Rasmussen M;
```

XX
DR WPI; 2000-365572/31.
DR P-PSDB; AAY71046.
XX
PT New alpha2M binding protein for generating a protective immune response
PT to group A streptococcus and purifying the binding protein -
XX
PS Claim 13; Page 65; 67pp; English.
XX
CC The patent discloses a new family of proteins termed GRAB (protein G
CC related alpha2M binding protein) from Streptococcus pyogenes which have
CC the ability to bind alpha2-macroglobulin (alpha2M) and show homology to
CC protein G of group G Streptococcus. GRAB protein and peptides derived
CC from it are used in vaccine compositions for generating a protective
CC immune response against group A Streptococcus. Antibodies against GRAB
CC are useful for treating Streptococcus pyogenes infections. The protein
CC is also useful for purifying alpha2M from a sample. The present sequence
CC is a DNA encoding partial GRAB protein from S. pyogenes strain KTL3.
CC The protein has alpha2M binding region and is useful in vaccine
CC composition.
XX
SQ Sequence 504 BP; 188 A; 97 C; 108 G; 111 T; 0 other;

Query Match 75.2%; Score 15.8; DB 21; Length 504;
Best Local Similarity 89.5%; Pred. No. 74;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 actgggatcaacagtatct 21
| |||| ||||||||||||
Db 12 agtgggttcaacagtatct 30

RESULT 6
AAX30352
ID AAX30352 standard; DNA; 697 BP.
XX
AC AAX30352;
XX
DT 14-MAY-1999 (first entry)
XX
DE DNA encoding a human secreted protein.
XX
KW Secreted protein; cancer; tumour; neurodegenerative disorder;
KW developmental abnormality; foetal deficiency; blood disorder;
KW CNS disorder; immune system disease; autoimmune disease; hepatic disease;
KW renal disease; diabetes; inflammation; allergy; ischemic shock;
KW Alzheimer's; cognitive disorder; schizophrenia; cardiovascular disorder;
KW prostate disease; asthma; osteoporosis; arthritis; ss.
XX
OS Homo sapiens.
XX
PN WO9907891-A1.
XX
PD 18-FEB-1999.
XX
PF 04-AUG-1998; 98WO-US16235.
XX
PR 19-AUG-1997; 97US-0056732.
PR 05-AUG-1997; 97US-0054798.
PR 05-AUG-1997; 97US-0054803.
PR 05-AUG-1997; 97US-0054804.
PR 05-AUG-1997; 97US-0054806.
PR 05-AUG-1997; 97US-0054807.
PR 05-AUG-1997; 97US-0054808.
PR 05-AUG-1997; 97US-0054809.
PR 05-AUG-1997; 97US-0055309.
PR 05-AUG-1997; 97US-0055310.
PR 05-AUG-1997; 97US-0055312.
PR 05-AUG-1997; 97US-0055386.
PR 05-AUG-1997; 97US-0055311.
PR 18-AUG-1997; 97US-0055970.
PR 18-AUG-1997; 97US-0055986.
PR 19-AUG-1997; 97US-0056365.
PR 19-AUG-1997; 97US-0056366.
PR 19-AUG-1997; 97US-0056557.
PR 19-AUG-1997; 97US-0056370.
PR 19-AUG-1997; 97US-0056371.
PR 19-AUG-1997; 97US-0056563.
PR 19-AUG-1997; 97US-0056731.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Brewer LA, Ebner R, Ferrie AM, Greene JM, Janat F, Ni J;
PI Olsen HS, Rosen CA, Ruben SM, Soppet DR, Young PE, Yu G;
XX
DR WPI; 1999-167452/14.
DR P-PSDB; AAY10831.
XX
PT New isolated human genes encoding secreted polypeptides - useful for
PT diagnosis and treatment of pathalogical diseases
XX
PS Claim 3; Page 247; 331pp; English.
XX
CC The specification describes secreted proteins and their corresponding
CC polynucleotides which are useful for preventing, treating or ameliorating
CC medical conditions, e.g. by protein or gene therapy. Pathological
CC conditions can also be diagnosed by determining the amount of the
CC secreted polypeptides in a sample or by determining the presence of
CC mutations in the polynucleotides. Specific uses are described for each
CC of the products, based on which tissues they are most highly
CC expressed in, and include developing products for the diagnosis or
CC treatment of cancer, tumours, neurodegenerative disorders, developmental
CC abnormalities and foetal deficiencies, blood disorders, CNS disorders,
CC diseases of the immune system, autoimmune diseases, hepatic and renal
CC disease, diabetes, inflammation, allergies, ischemic shock, Alzheimer's
CC and cognitive disorders, schizophrenia, cardiovascular disorders,
CC prostate diseases, asthma, disorders involving osteoclasts such as
CC osteoporosis, arthritis or malignancies, diseases of testes, lung or
CC thymus, digestive/endocrine disorders, infections and AIDS. The
CC polypeptides are also useful for identifying their binding partners.
XX
SQ Sequence 697 BP; 233 A; 121 C; 171 G; 169 T; 3 other;

Query Match 75.2%; Score 15.8; DB 20; Length 697;
Best Local Similarity 89.5%; Pred. No. 77;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 cactgggatcaacagtatc 20
||| |||| ||||||||||
Db 274 cacagggaacaacagtatc 292

RESULT 7
AAV30310
ID AAV30310 standard; DNA; 1073 BP.
XX
AC AAV30310;
XX
DT 28-SEP-1998 (first entry)
XX
DE Bacillus thuringiensis MIS-6 insecticidal toxin 196F3 DNA.
XX
KW Insecticide; pesticide; toxin; MIS-6; delta-endotoxin;
KW biological control; lepidopteran; coleopteran; ss.
XX
OS Bacillus thuringiensis strain PS196F3 (NRRL B-21872).
XX
FH Key Location/Qualifiers
FT CDS 2..1072
FT /*tag= a
XX
PN WO9818932-A2.
XX
PD 07-MAY-1998.

XX
PF 30-OCT-1997; 97WO-US19804.
XX
PR 30-OCT-1996; 96US-0029848.
XX
PA (MYCO) MYCOGEN CORP.
XX
PI Dullum CJ, Feitelson JS, Loewer D, Muller-Cohn J;
PI Narva KE, Schmeits JL, Schnepf HE, Schwab G, Stamp L;
PI Stockhoff BA;
XX
DR WPI; 1998-272226/24.
DR P-PSDB; AAW60227.
XX
PT Bacillus thuringiensis isolates - used for producing pesticidal
PT toxins and nucleotide sequences for control of lepidopterans and
PT coleopterans
XX
PS Claim 5; Page 90; 139pp; English.
XX
CC This DNA sequence encodes a novel soluble toxin (see AAW60227) of
CC Bacillus thuringiensis (B.t.) strain PS196F3 (NRRL B-21872). This
CC toxin belongs to the novel MIS-6 family of B.t. toxins that have
CC toxicity against non-mammalian pests. The novel DNA was obtained
CC by PCR amplification (see AAV30298-99) of total cellular genomic
CC DNA. It can used to produce recombinant hosts (preferably plant
CC or bacterial) that express the toxin, or as a PCR primer or
CC hybridisation probe for use in identifying and characterising
CC MIS-6 family toxin genes. Disclosed and claimed are novel B.t.
CC isolates and toxins (see AAW60218-32) which have activity against
CC lepidopteran and/or coleopteran pests, isolated genes, probes
CC and primers (see AAV30288-321 and AAT99734-87), and transformed host
CC cells. The invention provides 8 entirely new families of toxins,
CC including MIS-6, from B.t. isolates. MIS toxins are also useful
CC for their ability to form pores in cell membranes, and can be used
CC to facilitate entry of a second agent into a target cell.
XX
SQ Sequence 1073 BP; 380 A; 201 C; 215 G; 270 T; 7 other;

Query Match 75.2%; Score 15.8; DB 19; Length 1073;
Best Local Similarity 89.5%; Pred. No. 82;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 actgggatcaacagtatct 21
||||||| |||||||||
Db 838 actgggaaaaacagtatct 856

RESULT 8
AAV70397
ID AAV70397 standard; cDNA; 4351 BP.
XX
AC AAV70397;
XX
DT 10-FEB-1999 (first entry)
XX
DE LRP5 isoform 2 longest open reading frame (also isoform 4,5,6).
XX
KW LRP5; LDL-receptor related protein; LRP-3; IDDM; diagnosis;
KW endocytosis; insulin dependent diabetes mellitus; autoimmune disease;
KW glomerulonephritis; inflammation; viral infection; osteoporosis;
KW hypercholesterolemia; Alzheimer's disease; low density lipoprotein; ss.
XX
OS Homo sapiens.
XX
PN WO9846743-A1.
XX
PD 22-OCT-1998.
XX
PF 15-APR-1998; 98WO-GB01102.
XX
PR 05-JUN-1997; 97US-0048740.
PR 15-APR-1997; 97US-0043553.
XX
PA (MERI) MERCK & CO INC.
PA (WELL) WELLCOME TRUST LTD.
XX
PI Caskey CT, Cox RD, Gerhold D, Hammond H, Hess JW;
PI Hey P, Kawaguchi Y, Merriman TR, Metzker ML, Nakagawa Y;
PI Phillips MS, Todd JA, Twells RCJ;
XX
DR WPI; 1998-594573/50.
XX
PT New isolated LDL-receptor related protein - used to develop products
PT for treating, e.g. elevated triglyceride levels, diabetes,
PT autoimmune disorders, inflammation or Alzheimer's disease
XX
PS Claim 7; Fig 11b; 200pp; English.
XX
CC The present sequence represents LRP5 (low density lipoprotein (LDL)
CC receptor related protein, previously designated LRP-3) isoform 2 cDNA.
CC Nucleic acid molecules (NAMs) encoding LRP5 can be used for determining
CC if an individual is susceptible to insulin dependent diabetes mellitus
CC (IDDM). The NAMs or proteins can be used for reducing triglyceride levels
CC in the serum of an individual. Therapies that affect LRP5 may also be
CC useful in the treatment of autoimmune diseases such as
CC glomerulonephritis, diseases and disorders involving disruption of
CC endocytosis and/or antigen presentation, cytokine clearance and/or
CC inflammation, viral infection, pathogenic bacterial toxin contamination,
CC elevation of free fatty acids or hypercholesterolemia, type 2 diabetes,
CC osteoporosis, Alzheimer's disease and cardiovascular disease. Products
CC from the present invention can also be used for detection, diagnosis and
CC drug screening.
XX
SQ Sequence 4351 BP; 875 A; 1435 C; 1309 G; 732 T; 0 other;

Query Match 75.2%; Score 15.8; DB 19; Length 4351;
Best Local Similarity 89.5%; Pred. No. 1e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 cactgggatcaacagtatc 20
|||||||| |||||| |||
Db 1716 cactgggaccaacagaatc 1734

RESULT 9
AAV70395
ID AAV70395 standard; cDNA; 4843 BP.
XX
AC AAV70395;
XX
DT 10-FEB-1999 (first entry)
XX
DE LRP5 cDNA longest open reading frame.
XX
KW LRP5; LDL-receptor related protein; LRP-3; IDDM; diagnosis;
KW endocytosis; insulin dependent diabetes mellitus; autoimmune disease;
KW glomerulonephritis; inflammation; viral infection; osteoporosis;
KW hypercholesterolemia; Alzheimer's disease; low density lipoprotein; ss.
XX
OS Homo sapiens.
XX
PN WO9846743-A1.
XX
PD 22-OCT-1998.
XX
PF 15-APR-1998; 98WO-GB01102.
XX
PR 05-JUN-1997; 97US-0048740.
PR 15-APR-1997; 97US-0043553.
XX
PA (MERI) MERCK & CO INC.
PA (WELL) WELLCOME TRUST LTD.
XX

PI Caskey CT, Cox RD, Gerhold D, Hammond H, Hess JW;
PI Hey P, Kawaguchi Y, Merriman TR, Metzker ML, Nakagawa Y;
PI Phillips MS, Todd JA, Twells RCJ;
XX
DR WPI; 1998-594573/50.
XX
PT New isolated LDL-receptor related protein - used to develop products
PT for treating, e.g. elevated triglyceride levels, diabetes,
PT autoimmune disorders, inflammation or Alzheimer's disease
XX
PS Example 1; Fig 5b; 200pp; English.
XX
CC The present sequence represents the longest open reading frame of LRP5
CC (low density lipoprotein (LDL) receptor related protein, previously
CC designated LRP-3) cDNA from the present invention. Nucleic acid
CC molecules (NAMs) encoding LRP5 can be used for determining if an
CC individual is susceptible to insulin dependent diabetes mellitus (IDDM).
CC The NAMs or proteins can be used for reducing triglyceride levels
CC in the serum of an individual. Therapies that affect LRP5 may also be
CC useful in the treatment of autoimmune diseases such as
CC glomerulonephritis, diseases and disorders involving disruption of
CC endocytosis and/or antigen presentation, cytokine clearance and/or
CC inflammation, viral infection, pathogenic bacterial toxin contamination,
CC elevation of free fatty acids or hypercholesterolemia, type 2 diabetes,
CC osteoporosis, Alzheimer's disease and cardiovascular disease. Products
CC from the present invention can also be used for detection, diagnosis and
CC drug screening.
XX
SQ Sequence 4843 BP; 953 A; 1601 C; 1478 G; 811 T; 0 other;

Query Match 75.2%; Score 15.8; DB 19; Length 4843;
Best Local Similarity 89.5%; Pred. No. 1e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 cactgggatcaacagtatc 20
||||||||| |||||| |||
Db 2207 cactgggaccaacagaatc 2225

RESULT 10
AAV70398
ID AAV70398 standard; cDNA; 4915 BP.
XX
AC AAV70398;
XX
DT 10-FEB-1999 (first entry)
XX
DE LRP5 isoform 3 putative open reading frame.
XX
KW LRP5; LDL-receptor related protein; LRP-3; IDDM; diagnosis;
KW endocytosis; insulin dependent diabetes mellitus; autoimmune disease;
KW glomerulonephritis; inflammation; viral infection; osteoporosis;
KW hypercholesterolemia; Alzheimer's disease; low density lipoprotein; ss.
XX
OS Homo sapiens.
XX
PN WO9846743-A1.
XX
PD 22-OCT-1998.
XX
PF 15-APR-1998; 98WO-GB01102.
XX
PR 05-JUN-1997; 97US-0048740.
PR 15-APR-1997; 97US-0043553.
XX
PA (MERI) MERCK & CO INC.
PA (WELL) WELLCOME TRUST LTD.
XX
PI Caskey CT, Cox RD, Gerhold D, Hammond H, Hess JW;
PI Hey P, Kawaguchi Y, Merriman TR, Metzker ML, Nakagawa Y;
PI Phillips MS, Todd JA, Twells RCJ;
XX
DR WPI; 1998-594573/50.
XX
PT New isolated LDL-receptor related protein - used to develop products
PT for treating, e.g. elevated triglyceride levels, diabetes,
PT autoimmune disorders, inflammation or Alzheimer's disease
XX
PS Example 1; Fig 12c; 200pp; English.
XX
CC The present invention describes LRP5 (low density lipoprotein (LDL)
CC receptor related protein, previously designated LRP-3). The present
CC sequence represents the putative open reading frame of LRP5 isoform 3.
CC Nucleic acid molecules (NAMs) encoding LRP5 can be used for determining
CC if an individual is susceptible to insulin dependent diabetes mellitus
CC (IDDM). The NAMs or proteins can be used for reducing triglyceride levels
CC in the serum of an individual. Therapies that affect LRP5 may also be
CC useful in the treatment of autoimmune diseases such as
CC glomerulonephritis, diseases and disorders involving disruption of
CC endocytosis and/or antigen presentation, cytokine clearance and/or
CC inflammation, viral infection, pathogenic bacterial toxin contamination,
CC elevation of free fatty acids or hypercholesterolemia, type 2 diabetes,
CC osteoporosis, Alzheimer's disease and cardiovascular disease. Products
CC from the present invention can also be used for detection, diagnosis and
CC drug screening.
XX
SQ Sequence 4915 BP; 992 A; 1614 C; 1474 G; 835 T; 0 other;

Query Match 75.2%; Score 15.8; DB 19; Length 4915;
Best Local Similarity 89.5%; Pred. No. 1e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 cactgggatcaacagtatc 20
||||||||| |||||| |||
Db 2279 cactgggaccaacagaatc 2297

RESULT 11
AAV85549
ID AAV85549 standard; cDNA; 5022 BP.
XX
AC AAV85549;
XX
DT 10-FEB-1999 (first entry)
XX
DE LRP5 isoform 5 cDNA.
XX
KW LRP5; LDL-receptor related protein; LRP-3; IDDM; diagnosis;
KW endocytosis; insulin dependent diabetes mellitus; autoimmune disease;
KW glomerulonephritis; inflammation; viral infection; osteoporosis;
KW hypercholesterolemia; Alzheimer's disease; low density lipoprotein; ss.
XX
OS Homo sapiens.
XX
PN WO9846743-A1.
XX
PD 22-OCT-1998.
XX
PF 15-APR-1998; 98WO-GB01102.
XX
PR 05-JUN-1997; 97US-0048740.
PR 15-APR-1997; 97US-0043553.
XX
PA (MERI) MERCK & CO INC.
PA (WELL) WELLCOME TRUST LTD.
XX
PI Caskey CT, Cox RD, Gerhold D, Hammond H, Hess JW;
PI Hey P, Kawaguchi Y, Merriman TR, Metzker ML, Nakagawa Y;
PI Phillips MS, Todd JA, Twells RCJ;
XX
DR WPI; 1998-594573/50.
XX
PT New isolated LDL-receptor related protein - used to develop products
PT for treating, e.g. elevated triglyceride levels, diabetes,

PT autoimmune disorders, inflammation or Alzheimer's disease
XX
PS Claim 7; Fig 14; 200pp; English.
XX
CC The present invention describes LRP5 (low density lipoprotein (LDL)
CC receptor related protein, previously designated LRP-3). The present
CC sequence represents the LRP5 isoform 5 cDNA.
CC Nucleic acid molecules (NAMs) encoding LRP5 can be used for determining
CC if an individual is susceptible to insulin dependent diabetes mellitus
CC (IDDM). The NAMs or proteins can be used for reducing triglyceride levels
CC in the serum of an individual. Therapies that affect LRP5 may also be
CC useful in the treatment of autoimmune diseases such as
CC glomerulonephritis, diseases and disorders involving disruption of
CC endocytosis and/or antigen presentation, cytokine clearance and/or
CC inflammation, viral infection, pathogenic bacterial toxin contamination,
CC elevation of free fatty acids or hypercholesterolemia, type 2 diabetes,
CC osteoporosis, Alzheimer's disease and cardiovascular disease. Products
CC from the present invention can also be used for detection, diagnosis and
CC drug screening.
XX
SQ Sequence 5022 BP; 1036 A; 1606 C; 1503 G; 877 T; 0 other;

Query Match 75.2%; Score 15.8; DB 19; Length 5022;
Best Local Similarity 89.5%; Pred. No. 1e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 cactgggatcaacagtatc 20
||||||||| |||||| |||
Db 2178 cactgggaccaacagaatc 2196

RESULT 12
AAV70396
ID AAV70396 standard; cDNA; 5098 BP.
XX
AC AAV70396;
XX
DT 10-FEB-1999 (first entry)
XX
DE LRP5 isoform 1 cDNA.
XX
KW LRP5; LDL-receptor related protein; LRP-3; IDDM; diagnosis;
KW endocytosis; insulin dependent diabetes mellitus; autoimmune disease;
KW glomerulonephritis; inflammation; viral infection; osteoporosis;
KW hypercholesterolemia; Alzheimer's disease; low density lipoprotein; ss.
XX
OS Homo sapiens.
XX
PN WO9846743-A1.
XX
PD 22-OCT-1998.
XX
PF 15-APR-1998; 98WO-GB01102.
XX
PR 05-JUN-1997; 97US-0048740.
PR 15-APR-1997; 97US-0043553.
XX
PA (MERI) MERCK & CO INC.
PA (WELL) WELLCOME TRUST LTD.
XX
PI Caskey CT, Cox RD, Gerhold D, Hammond H, Hess JW;
PI Hey P, Kawaguchi Y, Merriman TR, Metzker ML, Nakagawa Y;
PI Phillips MS, Todd JA, Twells RCJ;
XX
DR WPI; 1998-594573/50.
XX
PT New isolated LDL-receptor related protein - used to develop products
PT for treating, e.g. elevated triglyceride levels, diabetes,
PT autoimmune disorders, inflammation or Alzheimer's disease
XX
PS Claim 3; Fig 5a; 200pp; English.
XX
CC The present sequence represents LRP5 (low density lipoprotein (LDL)
CC receptor related protein, previously designated LRP-3) isoform 1 cDNA.
CC Nucleic acid molecules (NAMs) encoding LRP5 can be used for determining
CC if an individual is susceptible to insulin dependent diabetes mellitus
CC (IDDM). The NAMs or proteins can be used for reducing triglyceride levels
CC in the serum of an individual. Therapies that affect LRP5 may also be
CC useful in the treatment of autoimmune diseases such as
CC glomerulonephritis, diseases and disorders involving disruption of
CC endocytosis and/or antigen presentation, cytokine clearance and/or
CC inflammation, viral infection, pathogenic bacterial toxin contamination,
CC elevation of free fatty acids or hypercholesterolemia, type 2 diabetes,
CC osteoporosis, Alzheimer's disease and cardiovascular disease. Products
CC from the present invention can also be used for detection, diagnosis and
CC drug screening.
XX
SQ Sequence 5098 BP; 1035 A; 1645 C; 1541 G; 877 T; 0 other;

Query Match 75.2%; Score 15.8; DB 19; Length 5098;
Best Local Similarity 89.5%; Pred. No. 1e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 cactgggatcaacagtatc 20
||||||||| |||||| |||
Db 2255 cactgggaccaacagaatc 2273

RESULT 13
AAV86019
ID AAV86019 standard; cDNA; 5125 BP.
XX
AC AAV86019;
XX
DT 10-FEB-1999 (first entry)
XX
DE Lrp5 isoform 3 cDNA.
XX
KW LRP5; LDL-receptor related protein; LRP-3; IDDM; diagnosis;
KW endocytosis; insulin dependent diabetes mellitus; autoimmune disease;
KW glomerulonephritis; inflammation; viral infection; osteoporosis;
KW hypercholesterolemia; Alzheimer's disease; low density lipoprotein; ss.
XX
OS Homo sapiens.
XX
PN WO9846743-A1.
XX
PD 22-OCT-1998.
XX
PF 15-APR-1998; 98WO-GB01102.
XX
PR 05-JUN-1997; 97US-0048740.
PR 15-APR-1997; 97US-0043553.
XX
PA (MERI) MERCK & CO INC.
PA (WELL) WELLCOME TRUST LTD.
XX
PI Caskey CT, Cox RD, Gerhold D, Hammond H, Hess JW;
PI Hey P, Kawaguchi Y, Merriman TR, Metzker ML, Nakagawa Y;
PI Phillips MS, Todd JA, Twells RCJ;
XX
DR WPI; 1998-594573/50.
XX
PT New isolated LDL-receptor related protein - used to develop products
PT for treating, e.g. elevated triglyceride levels, diabetes,
PT autoimmune disorders, inflammation or Alzheimer's disease
XX
PS Example 1; Fig 12a; 200pp; English.
XX
CC The present invention describes LRP5 (low density lipoprotein (LDL)
CC receptor related protein, previously designated LRP-3). The present
CC sequence represents the Lrp5 isoform 3 cDNA.
CC Nucleic acid molecules (NAMs) encoding LRP5 can be used for determining
CC if an individual is susceptible to insulin dependent diabetes mellitus

CC (IDDM). The NAMs or proteins can be used for reducing triglyceride levels
CC in the serum of an individual. Therapies that affect LRP5 may also be
CC useful in the treatment of autoimmune diseases such as
CC glomerulonephritis, diseases and disorders involving disruption of
CC endocytosis and/or antigen presentation, cytokine clearance and/or
CC inflammation, viral infection, pathogenic bacterial toxin contamination,
CC elevation of free fatty acids or hypercholesterolemia, type 2 diabetes,
CC osteoporosis, Alzheimer's disease and cardiovascular disease. Products
CC from the present invention can also be used for detection, diagnosis and
CC drug screening.
XX
SQ Sequence 5125 BP; 1069 A; 1640 C; 1518 G; 898 T; 0 other;

Query Match 75.2%; Score 15.8; DB 19; Length 5125;
Best Local Similarity 89.5%; Pred. No. 1e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 cactgggatcaacagtatc 20
||||||||| |||||| |||
Db 2282 cactgggaccaacagaatc 2300

RESULT 14
AAV85550
ID AAV85550 standard; cDNA; 5162 BP.
XX
AC AAV85550;
XX
DT 10-FEB-1999 (first entry)
XX
DE LRP5 isoform 6 cDNA.
XX
KW LRP5; LDL-receptor related protein; LRP-3; IDDM; diagnosis;
KW endocytosis; insulin dependent diabetes mellitus; autoimmune disease;
KW glomerulonephritis; inflammation; viral infection; osteoporosis;
KW hypercholesterolemia; Alzheimer's disease; low density lipoprotein; ss.
XX
OS Homo sapiens.
XX
PN WO9846743-A1.
XX
PD 22-OCT-1998.
XX
PF 15-APR-1998; 98WO-GB01102.
XX
PR 05-JUN-1997; 97US-0048740.
PR 15-APR-1997; 97US-0043553.
XX
PA (MERI) MERCK & CO INC.
PA (WELL) WELLCOME TRUST LTD.
XX
PI Caskey CT, Cox RD, Gerhold D, Hammond H, Hess JW;
PI Hey P, Kawaguchi Y, Merriman TR, Metzker ML, Nakagawa Y;
PI Phillips MS, Todd JA, Twells RCJ;
XX
DR WPI; 1998-594573/50.
XX
PT New isolated LDL-receptor related protein - used to develop products
PT for treating, e.g. elevated triglyceride levels, diabetes,
PT autoimmune disorders, inflammation or Alzheimer's disease
XX
PS Claim 7; Fig 15a; 200pp; English.
XX
CC The present invention describes LRP5 (low density lipoprotein (LDL)
CC receptor related protein, previously designated LRP-3). The present
CC sequence represents the LRP5 isoform 6 cDNA.
CC Nucleic acid molecules (NAMs) encoding LRP5 can be used for determining
CC if an individual is susceptible to insulin dependent diabetes mellitus
CC (IDDM). The NAMs or proteins can be used for reducing triglyceride levels
CC in the serum of an individual. Therapies that affect LRP5 may also be
CC useful in the treatment of autoimmune diseases such as
CC glomerulonephritis, diseases and disorders involving disruption of
CC endocytosis and/or antigen presentation, cytokine clearance and/or
CC inflammation, viral infection, pathogenic bacterial toxin contamination,
CC elevation of free fatty acids or hypercholesterolemia, type 2 diabetes,
CC osteoporosis, Alzheimer's disease and cardiovascular disease. Products
CC from the present invention can also be used for detection, diagnosis and
CC drug screening.
XX
SQ Sequence 5162 BP; 1062 A; 1665 C; 1527 G; 908 T; 0 other;

Query Match 75.2%; Score 15.8; DB 19; Length 5162;
Best Local Similarity 89.5%; Pred. No. 1e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 cactgggatcaacagtatc 20
||||||||| |||||| |||
Db 2319 cactgggaccaacagaatc 2337

RESULT 15
AAV85551
ID AAV85551 standard; cDNA; 5166 BP.
XX
AC AAV85551;
XX
DT 10-FEB-1999 (first entry)
XX
DE LRP5 isoform 2 cDNA.
XX
KW LRP5; LDL-receptor related protein; LRP-3; IDDM; diagnosis;
KW endocytosis; insulin dependent diabetes mellitus; autoimmune disease;
KW glomerulonephritis; inflammation; viral infection; osteoporosis;
KW hypercholesterolemia; Alzheimer's disease; low density lipoprotein; ss.
XX
OS Homo sapiens.
XX
PN WO9846743-A1.
XX
PD 22-OCT-1998.
XX
PF 15-APR-1998; 98WO-GB01102.
XX
PR 05-JUN-1997; 97US-0048740.
PR 15-APR-1997; 97US-0043553.
XX
PA (MERI) MERCK & CO INC.
PA (WELL) WELLCOME TRUST LTD.
XX
PI Caskey CT, Cox RD, Gerhold D, Hammond H, Hess JW;
PI Hey P, Kawaguchi Y, Merriman TR, Metzker ML, Nakagawa Y;
PI Phillips MS, Todd JA, Twells RCJ;
XX
DR WPI; 1998-594573/50.
XX
PT New isolated LDL-receptor related protein - used to develop products
PT for treating, e.g. elevated triglyceride levels, diabetes,
PT autoimmune disorders, inflammation or Alzheimer's disease
XX
PS Claim 8; Fig 11a; 200pp; English.
XX
CC The present invention describes LRP5 (low density lipoprotein (LDL)
CC receptor related protein, previously designated LRP-3). The present
CC sequence represents the LRP5 isoform 2 cDNA.
CC Nucleic acid molecules (NAMs) encoding LRP5 can be used for determining
CC if an individual is susceptible to insulin dependent diabetes mellitus
CC (IDDM). The NAMs or proteins can be used for reducing triglyceride levels
CC in the serum of an individual. Therapies that affect LRP5 may also be
CC useful in the treatment of autoimmune diseases such as
CC glomerulonephritis, diseases and disorders involving disruption of
CC endocytosis and/or antigen presentation, cytokine clearance and/or
CC inflammation, viral infection, pathogenic bacterial toxin contamination,
CC elevation of free fatty acids or hypercholesterolemia, type 2 diabetes,
CC osteoporosis, Alzheimer's disease and cardiovascular disease. Products

CC from the present invention can also be used for detection, diagnosis and
CC drug screening.
XX
SQ Sequence 5166 BP; 1082 A; 1654 C; 1530 G; 899 T; 1 other;

Query Match 75.2%; Score 15.8; DB 19; Length 5166;
Best Local Similarity 89.5%; Pred. No. 1e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy       2 cactgggatcaacagtatc 20
           ||||||||  ||||||  |||
Db    2333 cactgggaccaacagaatc 2351
```

Search completed: July 25, 2001, 05:23:04
Job time: 4677 sec

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: July 25, 2001, 05:16:42 ; Search time 1290.33 Seconds
(without alignments)
251.736 Million cell updates/sec

Title: US-09-142-095-2
Perfect score: 21
Sequence: 1 ccactgggatcaacagtatct 21

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues

Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_ba3:*
4: gb_in1:*
5: gb_in2:*
6: gb_in3:*
7: gb_om:*
8: gb_ov:*
9: gb_pat1:*
10: gb_pat2:*
11: gb_ph:*
12: gb_pl1:*
13: gb_pl2:*
14: gb_pl3:*
15: gb_pl4:*
16: em_ba1:*
17: em_ba2:*
18: em_fun:*
19: em_htgo_hum:*
20: em_htgo_inv:*
21: em_htgo_rod:*
22: em_htg_hum1:*
23: em_htg_hum2:*
24: em_htg_hum3:*
25: em_htg_hum4:*
26: em_htg_hum5:*
27: em_htg_hum6:*
28: em_htg_hum7:*
29: em_htg_hum8:*
30: em_htg_inv1:*
31: em_htg_inv2:*
32: em_htg_other:*
33: em_htg_rod:*
34: em_hum1:*
35: em_hum2:*
36: em_hum3:*
37: em_hum4:*
38: em_hum5:*
39: em_hum6:*
40: em_hum7:*
41: em_in:*
42: em_om:*
43: em_or:*
44: em_ov:*
45: em_pat:*
46: em_ph:*
47: em_pl:*
48: em_ro:*
49: em_sts:*
50: em_sy:*
51: em_un:*
52: em_vi:*
53: gb_sts1:*
54: gb_sts2:*
55: gb_sts3:*
56: gb_sy:*
57: gb_un:*
58: gb_vi1:*
59: gb_vi2:*
60: gb_htg1:*
61: gb_htg2:*
62: gb_htg3:*
63: gb_htg4:*
64: gb_htg5:*
65: gb_htg6:*
66: gb_htg7:*
67: gb_htg8:*
68: gb_htg9:*
69: gb_htg10:*
70: gb_htg11:*
71: gb_htg12:*
72: gb_htg13:*
73: gb_htg14:*
74: gb_htg15:*
75: gb_htg16:*
76: gb_htg17:*
77: gb_htg18:*
78: gb_htg19:*
79: gb_htg20:*
80: gb_htg21:*
81: gb_htg22:*
82: gb_htg23:*
83: gb_htg24:*
84: gb_htg25:*
85: gb_pr1:*
86: gb_pr2:*
87: gb_pr3:*
88: gb_pr4:*
89: gb_pr5:*
90: gb_pr6:*
91: gb_pr7:*
92: gb_pr8:*
93: gb_pr9:*
94: gb_ro1:*
95: gb_ro2:*
96: gb_in4:*
97: gb_pr10:*
98: em_ba3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		21	100.0	21	9	A65501	A65501 Sequence 2
c 2		21	100.0	292	89	AF357220	AF357220 Homo sapi
c 3		21	100.0	531	89	AF352795	AF352795 Homo sapi
c 4		21	100.0	541	89	AF180372	AF180372 Homo sapi
c 5		21	100.0	918	88	AF110194	AF110194 Homo sapi
c 6		21	100.0	1190	97	HUMUGT1A	M84125 Human bilir
c 7		21	100.0	2351	97	HUMHUGBR1	M57899 Human bilir
c 8		21	100.0	3341	91	D87674	D87674 Homo sapien

	9	21	100.0	68770	86	AC006985	AC006985 Homo sapi
	10	21	100.0	176619	70	AC026497	AC026497 Homo sapi
c	11	21	100.0	198872	89	AF297093	AF297093 Homo sapi
c	12	19.4	92.4	1602	91	AY029169	AY029169 Macaca mu
c	13	19.4	92.4	1647	88	AF104339	AF104339 Macaca fa
	14	17.8	84.8	104480	60	AC007905	AC007905 Homo sapi
c	15	17.8	84.8	187507	93	HSDJ137K2	AL049820 Human DNA
	16	17.4	82.9	12887	1	AF077006	AF077006 Helicobac
	17	17	81.0	163245	63	AC013699	AC013699 Homo sapi
c	18	17	81.0	167290	79	AL355679	AL355679 Homo sapi
	19	17	81.0	168656	90	AL359922	AL359922 Human DNA
	20	16.8	80.0	795	53	CNS0727X	AL425875 clone BA0
c	21	16.8	80.0	818	71	AC035580	AC035580 Giardia i
	22	16.8	80.0	900	72	AC054151	AC054151 Giardia i
c	23	16.8	80.0	999	53	CNS071O2	AL425160 clone BA0
	24	16.8	80.0	1721	7	AF232676	AF232676 Sus scrof
	25	16.8	80.0	17998	5	CEC33D3	Z49867 Caenorhabdi
c	26	16.8	80.0	33513	63	AC012901	AC012901 Drosophil
c	27	16.8	80.0	66168	6	DMSEG0007	AL033125 Drosophil
c	28	16.8	80.0	66857	79	AL354799_3	Continuation (4 of
	29	16.8	80.0	123926	76	AC079633	AC079633 Oryza sat
	30	16.8	80.0	148870	92	HS173A13	AL035688 Human DNA
c	31	16.8	80.0	151441	71	AC036158	AC036158 Homo sapi
c	32	16.8	80.0	161419	84	CNS01RHX	AL162632 Homo sapi
	33	16.8	80.0	162960	79	AL158062	AL158062 Homo sapi
	34	16.8	80.0	163749	79	AL157833	AL157833 Homo sapi
	35	16.8	80.0	173965	90	AL159977	AL159977 Human DNA
c	36	16.8	80.0	182007	82	AL590125	AL590125 Homo sapi
	37	16.8	80.0	193260	69	AC025598	AC025598 Homo sapi
	38	16.8	80.0	199648	72	AC060771	AC060771 Homo sapi
	39	16.8	80.0	201530	83	CNS01DXK	AL139354 Homo sapi
	40	16.8	80.0	294218	4	AE003430	AE003430 Drosophil
	41	16.4	78.1	6127	59	IBVCD61	Z69629 Infectious
	42	16.4	78.1	9081	58	AIBVCG	Z30541 Avian infec
	43	16.4	78.1	20500	59	IBAORFAB	M94356 Avian infec
	44	16.4	78.1	27608	59	IBACGB	M95169 Avian infec
c	45	16.4	78.1	87960	65	AC017906	AC017906 Drosophil

ALIGNMENTS

```
RESULT   1
A65501
LOCUS       A65501          21 bp    DNA             PAT       29-MAR-1999
DEFINITION  Sequence 2 from Patent WO9732042.
ACCESSION   A65501
VERSION     A65501.1  GI:4531236
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 21)
  AUTHORS   Burchell,B.
  TITLE     DRUG TRIAL ASSAY SYSTEM
  JOURNAL   Patent: WO 9732042-A 2 04-SEP-1997;
            UNIV DUNDEE (GB)
COMMENT     Other publication AU 2224197 19970916.
FEATURES             Location/Qualifiers
     source          1. .21
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT        6 a      6 c      4 g      5 t
ORIGIN
```

```
  Query Match             100.0%;  Score 21;  DB 9;  Length 21;
  Best Local Similarity   100.0%;  Pred. No. 0.68;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ccactgggatcaacagtatct 21
          |||||||||||||||||||||
Db      1 CCACTGGGATCAACAGTATCT 21
```

```
RESULT   2
AF357220/c
LOCUS       AF357220       292 bp    DNA             PRI       27-MAR-2001
DEFINITION  Homo sapiens bilirubin UDP-glucuronosyltransferase 1 (UGT1) gene,
            UGT1*1 allele, promoter and partial cds.
ACCESSION   AF357220
VERSION     AF357220.1  GI:13448828
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 292)
  AUTHORS   McKie,K., Addington,T., Nguyen,T.S., Glendenning,M., Kutlar,F. and
            Kutlar,A.
  TITLE     Detection of TATA box TA repeat region [6(TA)repeat] of human
            bilirubin UDP-glucuronosyltransferase 1-1 gene (UGT1*1) in an
            African American individual
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 292)
  AUTHORS   McKie,K., Addington,T., Nguyen,T.S., Glendenning,M., Kutlar,F. and
            Kutlar,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-MAR-2001) Medicine/Hemoglobin DNA Laboratory, Medical
            College of Georgia, 15th Street, Augusta, GA 30912, USA
FEATURES             Location/Qualifiers
     source          1. .292
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /map="2q37"
                     /cell_type="WBC"
                     /tissue_type="whole blood"
     gene            1. .>292
                     /gene="UGT1"
                     /allele="UGT1*1"
     promoter        1. .64
                     /gene="UGT1"
     variation       28. .39
                     /gene="UGT1"
                     /note="most common type of repeat; 6(TA)"
     TATA_signal     28. .42
                     /gene="UGT1"
     repeat_region   28. .39
                     /note="polymorphic region"
                     /rpt_type=tandem
                     /rpt_unit=ta
     mRNA            65. .>292
                     /gene="UGT1"
                     /product="bilirubin UDP-glucuronosyltransferase 1"
     CDS             81. .>292
                     /gene="UGT1"
                     /codon_start=1
                     /product="bilirubin UDP-glucuronosyltransferase 1"
                     /protein_id="AAK27223.1"
                     /db_xref="GI:13448829"
                     /translation="MAVESQGGRPLVLGLLLCVLGPVVSHAGKILLIPVDGSHWLSML
                     GAIQQLQQRGHEIVVLAPDASLYIRDG"
BASE COUNT       60 a     69 c     95 g     68 t
ORIGIN
```

```
  Query Match             100.0%;  Score 21;  DB 89;  Length 292;
  Best Local Similarity   100.0%;  Pred. No. 0.72;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ccactgggatcaacagtatct 21
          |||||||||||||||||||||
Db    186 CCACTGGGATCAACAGTATCT 166
```

```
RESULT   3
AF352795/c
LOCUS       AF352795      531 bp    DNA             PRI       10-APR-2001
DEFINITION  Homo sapiens bilirubin UDP-glucronosyltrasferase 1-1 (UGT1A1) gene,
            UGT1A1*1 allele, partial cds.
ACCESSION   AF352795
VERSION     AF352795.1  GI:13569708
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 531)
  AUTHORS   McKie,K., Kutlar,F., Glendenning,M. and Kutlar,A.
  TITLE     7(TA) repeat polymorphism of the TATA box of human bilirubin
            UDP-glucuronosyltransferase 1-1(UGT1A1*1) gene in a patient with
            sickle cell anemia + high bilirubinemia
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 531)
  AUTHORS   McKie,K., Kutlar,F., Glendenning,M. and Kutlar,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-FEB-2001) Medicine/Hemoglobin DNA Laboratory, Medical
            College of Georgia, 15th St., AC-1000, Augusta, GA 30912, USA
FEATURES             Location/Qualifiers
     source          1. .531
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /map="2q37"
                     /cell_type="white blood cell"
                     /tissue_type="blood"
     gene            <1. .>531
                     /gene="UGT1A1"
                     /note="UGT1"
                     /allele="UGT1A1*1"
     promoter        <1. .313
                     /gene="UGT1A1"
     TATA_signal     279. .295
                     /gene="UGT1A1"
     repeat_region   279. .>292
                     /note="repeat polymorphism compared to UGT1A1 sequence
                     presented in GenBank Accession Number AF180372; contains 7
                     ta repeats frequently detected in patients with high
                     bilirubinemia"
                     /rpt_type=tandem
                     /rpt_unit=ta
     mRNA            314. .>531
                     /gene="UGT1A1"
                     /product="bilirubin UDP-glucronosyltrasferase 1-1"
     CDS             334. .>531
                     /gene="UGT1A1"
                     /codon_start=1
                     /product="bilirubin UDP-glucronosyltrasferase 1-1"
                     /protein_id="AAK31204.1"
                     /db_xref="GI:13569709"
                     /translation="MAVESQGGRPLVLGLLLCVLGPVVSHAGKILLIPVDGSHWLSML
                     GAIQQLQQRGHEIVVLAPDASL"
BASE COUNT      120 a    121 c    137 g    153 t
ORIGIN


  Query Match           100.0%;  Score 21;  DB 89;  Length 531;
  Best Local Similarity  100.0%;  Pred. No. 0.73;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ccactgggatcaacagtatct 21
          |||||||||||||||||||||
Db    439 CCACTGGGATCAACAGTATCT 419


RESULT   4
AF180372/c
LOCUS       AF180372      541 bp    DNA             PRI       05-OCT-1999
```

```
DEFINITION  Homo sapiens bilirubin UDP-glucuronosyltransferase 1-1 (UGT1) gene,
            UGT1*1 allele, partial cds.
ACCESSION   AF180372
VERSION     AF180372.1  GI:6010649
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 541)
  AUTHORS   Kutlar,F., Sromek,E., Leithner,C., Nechtman,J. and Kutlar,A.
  TITLE     Detection of the TATA box polymorphism of the human bilirubin
            UDP-glucuronosyltransferase 1-1 gene (UGT1*1) in a patient with
            sickle cell anemia
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 541)
  AUTHORS   Kutlar,F., Sromek,E., Leithner,C., Nechtman,J. and Kutlar,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-AUG-1999) Medicine, Hematology/Oncology-Sickle Cell
            Center, Medical College of Georgia, 15th Street, AC-1000, Augusta,
            GA 30912, USA
FEATURES             Location/Qualifiers
     source          1. .541
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /map="2q37"
                     /cell_type="white blood cells"
                     /tissue_type="whole blood"
     gene            1. .>541
                     /gene="UGT1"
                     /note="GNT1; UGT1A1"
                     /allele="UGT1*1"
     promoter        1. .313
                     /gene="UGT1"
     variation       226
                     /gene="UGT1"
                     /replace="t"
     repeat_region   279. .288
                     /note="polymorphic region"
                     /rpt_type=tandem
                     /rpt_unit=ta
     TATA_signal     279. .291
                     /gene="UGT1"
                     /note="Ritter,J.K., et al., 1992, J. Biol. Chem.,
                     267:3257-3261"
     mRNA            314. .>541
                     /gene="UGT1"
                     /product="bilirubin UDP-glucuronosyltransferase 1-1"
     CDS             330. .>541
                     /gene="UGT1"
                     /note="UDP glycosyltransferase 1"
                     /codon_start=1
                     /product="bilirubin UDP-glucuronosyltransferase 1-1"
                     /protein_id="AAF01205.1"
                     /db_xref="GI:6010650"
                     /translation="MAVESQGGRPLVLGLLLCVLGPVVSHAGKILLIPVDGSHWLSML
                     GAIQQLQQRGHEIVVLAPDASLYIRDG"
BASE COUNT      123 a    124 c    141 g    153 t
ORIGIN


  Query Match           100.0%;  Score 21;  DB 89;  Length 541;
  Best Local Similarity  100.0%;  Pred. No. 0.73;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ccactgggatcaacagtatct 21
          |||||||||||||||||||||
Db    435 CCACTGGGATCAACAGTATCT 415


RESULT   5
AF110194/c
```

```
LOCUS       AF110194      918 bp    DNA             PRI       02-JAN-2001
DEFINITION  Homo sapiens chromosome 2 UDP-glucuronosyltransferase (UGT1A1)
            gene, UGT1A1*33 allele, partial cds.
ACCESSION   AF110194
VERSION     AF110194.1  GI:12002134
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 918)
  AUTHORS   Guillemette,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-DEC-1998) Center for Cancer Research, MIT, 77
            Massachusetts Avenue, E17-540, Cambridge, MA 02139, USA
FEATURES             Location/Qualifiers
     source          1. .918
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="2"
     mRNA            <38. .>918
                     /gene="UGT1A1"
                     /product="UDP-glucuronosyltransferase"
     gene            <38. .>918
                     /gene="UGT1A1"
                     /note="22. .918; 1A1 variant allele; L233R"
                     /allele="UGT1A1*33"
     CDS             38. .>918
                     /gene="UGT1A1"
                     /codon_start=1
                     /product="UDP-glucuronosyltransferase"
                     /protein_id="AAG43197.1"
                     /db_xref="GI:12002135"
                     /translation="MAVESQGGRPLVLGLLLCVLGPVVSHAGKILLIPVDGSHWLSML
                     GAIQQLQQRGHEIVVLAPDASLYIRDGAFYTLKTYPVPFQREDVKESFVSLGHNVFEN
                     DSFLQRVIKTYKKIKKDSAMLLSGCSHLLHNKELMASLAESSFDVMLTDPFLPCSPIV
                     AQYLSLPTVFFLHALPCSLEFEATQCPNPFSYVPRPLSSHSDHMTFLQRVKNMLIAFS
                     QNFLCDVVYSPYATRASEFLQREVTVQDLLSSASVWLFRSDFVKDYPRPIMPNMVFVG
                     GINCLHQNPLSQVCIGVG"
BASE COUNT      196 a    245 c    235 g    242 t
ORIGIN
```

```
  Query Match           100.0%;  Score 21;  DB 88;  Length 918;
  Best Local Similarity  100.0%;  Pred. No. 0.74;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ccactgggatcaacagtatct 21
           |||||||||||||||||||||
Db     143 CCACTGGGATCAACAGTATCT 123
```

```
RESULT   6
HUMUGT1A/c
LOCUS       HUMUGT1A     1190 bp    DNA             PRI       14-JAN-1995
DEFINITION  Human bilirubin UDP-glucuronosyltransferase (UGT1A) gene isozyme
            exon 1.
ACCESSION   M84125
VERSION     M84125.1  GI:340131
KEYWORDS    bilirubin UDP-glucuronosyltransferase; isozyme.
SOURCE      Homo sapiens (tissue library: cosmid) liver DNA.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1190)
  AUTHORS   Ritter,J.K., Crawford,J.M. and Owens,I.S.
  TITLE     Cloning of two human liver bilirubin UDP-glucuronosyltransferase
            cDNAs with expression in COS-1 cells
  JOURNAL   J. Biol. Chem. 266 (2), 1043-1047 (1991)
  MEDLINE   91093210
REFERENCE   2  (bases 1 to 1190)
  AUTHORS   Ritter,J.K., Chen,F., Sheen,Y.Y., Tran,H.M., Kimura,S.,
            Yeatman,M.T. and Owens,I.S.
  TITLE     A novel complex locus UGT1 encodes human bilirubin, phenol, and
            other UDP-glucuronosyltransferase isozymes with identical carboxyl
            termini
  JOURNAL   J. Biol. Chem. 267 (5), 3257-3261 (1992)
  MEDLINE   92147680
FEATURES             Location/Qualifiers
     source          1. .1190
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /tissue_type="liver"
                     /tissue_lib="cosmid"
                     /lab_host="XL1-Blue"
     gene            32. .46
                     /gene="UTG1A"
     TATA_signal     32. .46
                     /gene="UTG1A"
     exon            69. .948
                     /gene="UGT1A"
                     /number=1
     gene            69. .1190
                     /gene="UGT1A"
     CDS             85. .>948
                     /gene="UGT1A"
                     /codon_start=1
                     /product="bilirubin UDP-glucuronosyltransferase"
                     /protein_id="AAA61248.1"
                     /db_xref="GI:340132"
                     /translation="MAVESQGGRPLVLGLLLCVLGPVVSHAGKILLIPVDGSHWLSML
                     GAIQQLQQRGHEIVVLAPDASLYIRDGAFYTLKTYPVPFQREDVKESFVSLGHNVFEN
                     DSFLQRVIKTYKKIKKDSAMLLSGCSHLLHNKELMASLAESSFDVMLTDPFLPCSPIV
                     AQYLSLPTVFFLHALPCSLEFEATQCPNPFSYVPRPLSSHSDHMTFLQRVKNMLIAFS
                     QNFLCDVVYSPYATLASEFLQREVTVQDLLSSASVWLFRSDFVKDYPRPIMPNMVFVG
                     GINCLHQNPLSQ"
     intron          949. .1190
                     /partial
                     /gene="UGT1A"
                     /note="does not fit consensus"
                     /number=1
BASE COUNT      261 a    290 c    286 g    353 t
ORIGIN
```

```
  Query Match           100.0%;  Score 21;  DB 97;  Length 1190;
  Best Local Similarity  100.0%;  Pred. No. 0.74;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ccactgggatcaacagtatct 21
           |||||||||||||||||||||
Db     190 CCACTGGGATCAACAGTATCT 170
```

```
RESULT   7
HUMHUGBR1/c
LOCUS       HUMHUGBR1    2351 bp    mRNA            PRI       07-MAR-1995
DEFINITION  Human bilirubin UDP-glucuronosyltransferase isozyme 1 mRNA,
            complete cds.
ACCESSION   M57899
VERSION     M57899.1  GI:184472
KEYWORDS    UDP-glucuronosyltransferase; bilirubin UDP-glucuronosyltransferase.
SOURCE      Human adult female liver, cDNA to mRNA, clones Z6, Z11 and Z6MB2.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2351)
  AUTHORS   Ritter,J.K., Crawford,J.M. and Owens,I.S.
  TITLE     Cloning of two human liver bilirubin UDP-glucuronosyltransferase
            cDNAs with expression in COS-1 cells
  JOURNAL   J. Biol. Chem. 266 (2), 1043-1047 (1991)
  MEDLINE   91093210
FEATURES             Location/Qualifiers
     source          1. .2351
                     /organism="Homo sapiens"
                     /isolate="AK"
```

```
                     /db_xref="taxon:9606"
                     /clone="Z6"
                     /dev_stage="adult"
                     /sex="female"
                     /tissue_type="liver"
                     /tissue_lib="lambda-ZAP"
                     /map="2"
     gene            16. .1617
                     /gene="UGT1"
     CDS             16. .1617
                     /gene="UGT1"
                     /standard_name="bilirubin UDP-glucuronosyltransferase
                     isozyme 1"
                     /EC_number="2.4.1.17"
                     /codon_start=1
                     /db_xref="GDB:G00-120-007"
                     /product="UDP-glucuronosyltransferase 1"
                     /protein_id="AAA63195.1"
                     /db_xref="GI:184473"
                     /translation="MAVESQGGRPLVLGLLLCVLGPVVSHAGKILLIPVDGSHWLSML
                     GAIQQLQQRGHEIVVLAPDASLYIRDGAFYTLKTYPVPFQREDVKESFVSLGHNVFEN
                     DSFLQRVIKTYKKIKKDSAMLLSGCSHLLHNKELMASLAESSFDVMLTDPFLPCSPIV
                     AQYLSLPTVFFLHALPCSLEFEATQCPNPFSYVPRPLSSHSDHMTFLQRVKNMLIAFS
                     QNFLCDVVYSPYATLASEFLQREVTVQDLLSSASVWLFRSDFVKDYPRPIMPNMVFVG
                     GINCLHQNPLSQEFEAYINASGEHGIVVFSLGSMVSEIPEKKAMAIADALGKIPQTVL
                     WRYTGTRPSNLANNTILVKWLPQNDLLGHPMTRAFITHAGSHGVYESICNGVPMVMMP
                     LFGDQMDNAKRMETKGAGVTLNVLEMTSEDLENALKAVINDKSYKENIMRLSSLHKDR
                     PVEPLDLAVFWVEFVMRHKGAPHLRPAAHDLTWYQYHSLDVIGFLLAVVLTVAFITFK
                     CCAYGYRKCLGKKGRVKKAHKSKTH"
     sig_peptide     43. .75
                     /gene="UGT1"
                     /note="G00-120-007"
     mat_peptide     76. .1614
                     /gene="UGT1"
                     /standard_name="bilirubin UDP-glucuronosyltransferase
                     isozyme 1"
                     /EC_number="2.4.1.17"
                     /note="G00-120-007"
                     /product="UDP-glucuronosyltransferase 1"
BASE COUNT      601 a    540 c    557 g    653 t
ORIGIN
```

```
  Query Match             100.0%;  Score 21;  DB 97;  Length 2351;
  Best Local Similarity   100.0%;  Pred. No. 0.76;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ccactgggatcaacagtatct 21
          |||||||||||||||||||||
Db    121 CCACTGGGATCAACAGTATCT 101
```

```
RESULT   8
D87674/c
LOCUS       D87674       3341 bp    DNA             PRI       14-APR-2000
DEFINITION  Homo sapiens gene for bilirubin UDP-glucuronosyltransferase 1,
            promoter region and partial cds.
ACCESSION   D87674
VERSION     D87674.1  GI:3059176
KEYWORDS    bilirubin UDP-glucuronosyltransferase 1.
SOURCE      Homo sapiens DNA.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Ueyama,H., Koiwai,O., Soeda,Y., Sato,H., Satoh,Y., Ohkubo,I. and
            Doida,Y.
  TITLE     Analysis of the promoter of human bilirubin
            UDP-glucuronosyltransferase gene (UGT1*1) in relevance to Gilbert's
            syndrome
  JOURNAL   Hepatol. Res. 9, 152-163 (1997)
REFERENCE   2  (bases 1 to 3341)
  AUTHORS   Ueyama,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-SEP-1996) to the DDBJ/EMBL/GenBank databases. Hisao
            Ueyama, Shiga University of Medical Science, Department of Medical
            Biochemistry; Seta, Otsu, Shiga 520-21, Japan (Tel:077-548-2162,
            Fax:077-548-2164)
COMMENT     Sequence updated (08-Jan-1997) by: Hisao Ueyama.
FEATURES             Location/Qualifiers
     source          1. .3341
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="2q37"
     misc_feature    2538. .2544
                     /note="XRE"
     misc_feature    2606. .2610
                     /note="XRE"
     misc_feature    3088. .3097
                     /note="E-box"
     misc_feature    3101. .3113
                     /note="HNF-1 site"
     CAAT_signal     3125. .3129
     TATA_signal     3149. .3153
     exon            3177. .3341
                     /number=1
     gene            3192. .3341
                     /gene="UGT1*1"
     CDS             3192. .>3341
                     /gene="UGT1*1"
                     /codon_start=1
                     /product="bilirubin UDP-glucuronosyltransferase 1"
                     /protein_id="BAA25600.1"
                     /db_xref="GI:3059177"
                     /translation="MAVESQGGRPLVLGLLLCVLGPVVSHAGKILLIPVDGSHWLSML
                     GAIQQL"
BASE COUNT      893 a    695 c    803 g    950 t
ORIGIN
```

```
  Query Match             100.0%;  Score 21;  DB 91;  Length 3341;
  Best Local Similarity   100.0%;  Pred. No. 0.76;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ccactgggatcaacagtatct 21
          |||||||||||||||||||||
Db   3297 CCACTGGGATCAACAGTATCT 3277
```

```
RESULT   9
AC006985
LOCUS       AC006985    68770 bp    DNA             PRI       21-DEC-1999
DEFINITION  Homo sapiens BAC clone RP11-154L24 from 2, complete sequence.
ACCESSION   AC006985
VERSION     AC006985.2  GI:5732165
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 68770)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
  MEDLINE   99063792
REFERENCE   2  (bases 1 to 68770)
  AUTHORS   Gattung,S., Stoneking,T. and Davidson,T.
  TITLE     The sequence of Homo sapiens BAC clone RP11-154L24
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 68770)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-MAR-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 68770)
```

AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (13-AUG-1999) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE 5 (bases 1 to 68770)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (22-OCT-1999) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE 6 (bases 1 to 68770)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (21-DEC-1999) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT On Aug 13, 1999 this sequence version replaced gi:4337256.
---------- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@watson.wustl.edu
---------- Summary Statistics
Center project name: H_NH0154L24
----------.

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E., Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (http://bacpac.med.buffalo.edu)
VECTOR: pBACe3.6
NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is RP11-332L11, 200 bp overlap. Actual start of this clone is at base position 86134 of RP11-332L11; actual end is at base position 68770 of RP11-154L24.

The clone RP11-154L24 contains a tandem repeat from base positions 38234 to 39039, this region contains some low quality data. The assembly is consistent with the restriction digest information.

FEATURES Location/Qualifiers
source 1. .68770
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/map="2"
/clone="RP11-154L24"
/clone_lib="RPCI-11"
repeat_region 3. .192
/rpt_family="L1"
repeat_region 221. .528
/rpt_family="Alu"
repeat_region 529. .1634
/rpt_family="L1"
repeat_region 1658. .1878
/rpt_family="MIR"
repeat_region 2079. .2271
/rpt_family="MER1_type"
repeat_region 2276. .2336
/rpt_family="MER1_type"
repeat_region 4683. .4806
/rpt_family="MIR"
repeat_region 5131. .5389
/rpt_family="MaLR"
repeat_region 8738. .8871
/rpt_family="MIR"
repeat_region 9092. .9371
/rpt_family="AcHobo"
repeat_region 9967. .10582
/rpt_family="L1"
repeat_region 10576. .11200
/rpt_family="L1"
repeat_region 11196. .11469
/rpt_family="L1"
repeat_region 11540. .11654
/rpt_family="L1"
repeat_region 12108. .12523
/rpt_family="MaLR"
repeat_region 13113. .13149
/rpt_family="MER1_type"
repeat_region 13150. .13171
/rpt_family="(CAG)n"
repeat_region 13172. .13286
/rpt_family="MER1_type"
repeat_region 13338. .13497
/rpt_family="MER1_type"
repeat_region 14727. .14848
/rpt_family="MIR"
repeat_region 15890. .16244
/rpt_family="MaLR"
repeat_region 16380. .16723
/rpt_family="L1"
repeat_region 17483. .17560
/rpt_family="L2"
repeat_region 17655. .17747
/rpt_family="L2"
repeat_region 18220. .18337
/rpt_family="MER1_type"
repeat_region 18770. .18816
/rpt_family="(CA)n"
repeat_region 19103. .19121
/rpt_family="AT_rich"
repeat_region 19122. .19636
/rpt_family="L1"
repeat_region 19637. .19662
/rpt_family="AT_rich"
repeat_region 19825. .20014
/rpt_family="L1"
repeat_region 20034. .20098
/rpt_family="MIR"
repeat_region 20254. .21067
/rpt_family="L1"
repeat_region 21141. .21182
/rpt_family="A-rich"
repeat_region 24649. .25018
/rpt_family="Retroviral"
repeat_region 25258. .25447
/rpt_family="MIR"
repeat_region 25777. .25824
/rpt_family="(CA)n"
repeat_region 28020. .28178
/rpt_family="MIR"
repeat_region 28178. .28222
/rpt_family="L2"

```
     repeat_region   28503. .28558
                     /rpt_family="L1"
     repeat_region   28558. .28741
                     /rpt_family="L1"
     repeat_region   28970. .29064
                     /rpt_family="L1"
     repeat_region   29276. .29434
                     /rpt_family="MER1_type"
     repeat_region   29554. .29712
                     /rpt_family="MER1_type?"
     repeat_region   30867. .31015
                     /rpt_family="MER1_type"
     repeat_region   31268. .31682
                     /rpt_family="MaLR"
     repeat_region   32195. .32390
                     /rpt_family="MIR"
     repeat_region   33607. .33760
                     /rpt_family="L2"
     repeat_region   33835. .34127
                     /rpt_family="Alu"
     repeat_region   34167. .34283
                     /rpt_family="GA-rich"
     repeat_region   34806. .34976
                     /rpt_family="L2"
     repeat_region   35247. .35719
                     /rpt_family="L2"
     repeat_region   35720. .35762
                     /rpt_family="GA-rich"
     repeat_region   35763. .35909

  Query Match           100.0%;  Score 21;  DB 86;  Length 68770;
  Best Local Similarity 100.0%;  Pred. No. 0.81;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ccactgggatcaacagtatct 21
           |||||||||||||||||||||
Db   66050 CCACTGGGATCAACAGTATCT 66070


RESULT  10
AC026497
LOCUS       AC026497   176619 bp    DNA             HTG       22-MAR-2000
DEFINITION  Homo sapiens chromosome 11 clone RP11-689A10 map 11, *** SEQUENCING
            IN PROGRESS ***, 43 unordered pieces.
ACCESSION   AC026497
VERSION     AC026497.1  GI:7283186
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 176619)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 11, clone RP11-689A10
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 176619)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-MAR-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            ------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            ------------- Project Information
                Center project name: L6277
                Center clone name: 689_A_10
            -------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 43 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1333: contig of 1333 bp in length
            *     1334  1433: gap of      100 bp
            *     1434     2787: contig of 1354 bp in length
            *     2788  2887: gap of      100 bp
            *     2888     3998: contig of 1111 bp in length
            *     3999  4098: gap of      100 bp
            *     4099     5777: contig of 1679 bp in length
            *     5778  5877: gap of      100 bp
            *     5878     7685: contig of 1808 bp in length
            *     7686  7785: gap of      100 bp
            *     7786     8176: contig of 391 bp in length
            *     8177  8276: gap of      100 bp
            *     8277     9926: contig of 1650 bp in length
            *     9927 10026: gap of      100 bp
            *    10027    11499: contig of 1473 bp in length
            *    11500 11599: gap of      100 bp
            *    11600    13616: contig of 2017 bp in length
            *    13617 13716: gap of      100 bp
            *    13717    16073: contig of 2357 bp in length
            *    16074 16173: gap of      100 bp
            *    16174    18510: contig of 2337 bp in length
            *    18511 18610: gap of      100 bp
            *    18611    20495: contig of 1885 bp in length
            *    20496 20595: gap of      100 bp
            *    20596    23548: contig of 2953 bp in length
            *    23549 23648: gap of      100 bp
            *    23649    25910: contig of 2262 bp in length
            *    25911 26010: gap of      100 bp
            *    26011    28060: contig of 2050 bp in length
            *    28061 28160: gap of      100 bp
            *    28161    30134: contig of 1974 bp in length
            *    30135 30234: gap of      100 bp
            *    30235    32646: contig of 2412 bp in length
            *    32647 32746: gap of      100 bp
            *    32747    35096: contig of 2350 bp in length
            *    35097 35196: gap of      100 bp
            *    35197    37696: contig of 2500 bp in length
            *    37697 37796: gap of      100 bp
            *    37797    40539: contig of 2743 bp in length
            *    40540 40639: gap of      100 bp
            *    40640    45518: contig of 4879 bp in length
            *    45519 45618: gap of      100 bp
            *    45619    48728: contig of 3110 bp in length
            *    48729 48828: gap of      100 bp
            *    48829    52099: contig of 3271 bp in length
            *    52100 52199: gap of      100 bp
```

```
            *    52200    56747: contig of 4548 bp in length
            *    56748 56847: gap of      100 bp
            *    56848    60467: contig of 3620 bp in length
            *    60468 60567: gap of      100 bp
            *    60568    64947: contig of 4380 bp in length
            *    64948 65047: gap of      100 bp
            *    65048    69431: contig of 4384 bp in length
            *    69432 69531: gap of      100 bp
            *    69532    74167: contig of 4636 bp in length
            *    74168 74267: gap of      100 bp
            *    74268    78188: contig of 3921 bp in length
            *    78189 78288: gap of      100 bp
            *    78289    81664: contig of 3376 bp in length
            *    81665 81764: gap of      100 bp
            *    81765    85169: contig of 3405 bp in length
            *    85170 85269: gap of      100 bp
            *    85270    90445: contig of 5176 bp in length
            *    90446 90545: gap of      100 bp
            *    90546    95008: contig of 4463 bp in length
            *    95009 95108: gap of      100 bp
            *    95109   100587: contig of 5479 bp in length
            *   100588 100687: gap of      100 bp
            *   100688   105884: contig of 5197 bp in length
            *   105885 105984: gap of      100 bp
            *   105985   111486: contig of 5502 bp in length
            *   111487 111586: gap of      100 bp
            *   111587   117781: contig of 6195 bp in length
            *   117782 117881: gap of      100 bp
            *   117882   125343: contig of 7462 bp in length
            *   125344 125443: gap of      100 bp
            *   125444   132508: contig of 7065 bp in length
            *   132509 132608: gap of      100 bp
            *   132609   141190: contig of 8582 bp in length
            *   141191 141290: gap of      100 bp
            *   141291   152556: contig of 11266 bp in length
            *   152557 152656: gap of      100 bp
            *   152657   163590: contig of 10934 bp in length
            *   163591 163690: gap of      100 bp
            *   163691   176619: contig of 12929 bp in length.
FEATURES             Location/Qualifiers
     source          1. .176619
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11"
                     /clone="RP11-689A10"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1333
                     /note="assembly_fragment"
     misc_feature    1434. .2787
                     /note="assembly_fragment"
     misc_feature    2888. .3998
                     /note="assembly_fragment"
     misc_feature    4099. .5777
                     /note="assembly_fragment"
     misc_feature    5878. .7685
                     /note="assembly_fragment"
     misc_feature    7786. .8176
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:right"
     misc_feature    8277. .9926
                     /note="assembly_fragment"
     misc_feature    10027. .11499
                     /note="assembly_fragment"
     misc_feature    11600. .13616
                     /note="assembly_fragment"
     misc_feature    13717. .16073
                     /note="assembly_fragment"
     misc_feature    16174. .18510
                     /note="assembly_fragment"
     misc_feature    18611. .20495
                     /note="assembly_fragment"
     misc_feature    20596. .23548
                     /note="assembly_fragment"
     misc_feature    23649. .25910
                     /note="assembly_fragment"
     misc_feature    26011. .28060
                     /note="assembly_fragment"
     misc_feature    28161. .30134
                     /note="assembly_fragment"
     misc_feature    30235. .32646
                     /note="assembly_fragment"
     misc_feature    32747. .35096
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
     misc_feature    35197. .37696
                     /note="assembly_fragment"
     misc_feature    37797. .40539
                     /note="assembly_fragment"
     misc_feature    40640. .45518
                     /note="assembly_fragment"
     misc_feature    45619. .48728
                     /note="assembly_fragment"

  Query Match             100.0%;  Score 21;  DB 70;  Length 176619;
  Best Local Similarity   100.0%;  Pred. No. 0.83;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ccactgggatcaacagtatct 21
           |||||||||||||||||||||
Db   89296 CCACTGGGATCAACAGTATCT 89316


RESULT  11
AF297093/c
LOCUS       AF297093   198872 bp    DNA             PRI       07-NOV-2000
DEFINITION  Homo sapiens UGT1 gene locus, complete sequence.
ACCESSION   AF297093
VERSION     AF297093.1  GI:11118740
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 198872)
  AUTHORS   Owens,I.S., Gong,Q., Cho,J.W., Huang,T., Potter,C., Gholami,N.,
            Basu,N.K., Kubota,S., Carvalho,S. and Pennington,M.W.
  TITLE     Thirteen UDP glucuronosyltransferase genes encoded at the human
            UGT1 locus
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 198872)
  AUTHORS   Owens,I.S., Gong,Q., Cho,J.W., Huang,T., Potter,C., Gholami,N.,
            Basu,N.K., Kubota,S., Carvalho,S. and Pennington,M.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-AUG-2000) Heritable Disorders Branch, NIH-NICHD, 9000
            Rockville Pike, Bethesda, MD 20892, USA
FEATURES             Location/Qualifiers
     source          1. .198872
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     gene            1989. .2841
                     /gene="UGT1A12p"
                     /note="UDP glucuronosyltransferase 1A12"
                     /pseudo
     CDS             1989. .2841
                     /gene="UGT1A12p"
                     /codon_start=1
                     /pseudo
     gene            19970. .20956
                     /gene="UGT1A11p"
                     /note="UDP glucuronosyltransferase 1A11"
                     /pseudo
     TATA_signal     19970. .19978
     CDS             20105. .20956
```

```
                     /gene="UGT1A11p"
                     /codon_start=1
                     /pseudo
     TATA_signal     34133. .34149
                     /gene="UGT1A8"
     gene            34133. .>187313
                     /gene="UGT1A8"
     mRNA            join(<34264. .35118,181788. .181919,182603. .182690,
                     182974. .183193,187016. .>187313)
                     /gene="UGT1A8"
                     /product="UDP glucuronosyltransferase 1A8"
     CDS             join(34264. .35118,181788. .181919,182603. .182690,
                     182974. .183193,187016. .187313)
                     /gene="UGT1A8"
                     /codon_start=1
                     /product="UDP glucuronosyltransferase 1A8"
                     /protein_id="AAG30416.1"
                     /db_xref="GI:11118741"
                     /translation="MARTGWTSPIPLCVSLLLTCGFAEAGKLLVVPMDGSHWFTMQSV
                     VEKLILRGHEVVVVMPEVSWQLGKSLNCTVKTYSTSYTLEDLDREFMDFADAQWKAQV
                     RSLFSLFLSSSNGFFNLFFSHCRSLFNDRKLVEYLKESSFDAVFLDPFDACGLIVAKY
                     FSLPSVVFARGIACHYLEEGAQCPAPLSYVPRILLGFSDAMTFKERVRNHIMHLEEHL
                     FCQYFSKNALEIASEILQTPVTAYDLYSHTSIWLLRTDFVLDYPKPVMPNMIFIGGIN
                     CHQGKPLPMEFEAYINASGEHGIVVFSLGSMVSEIPEKKAMAIADALGKIPQTVLWRY
                     TGTRPSNLANNTILVKWLPQNDLLGHPMTRAFITHAGSHGVYESICNGVPMVMMPLFG
                     DQMDNAKRMETKGAGVTLNVLEMTSEDLENALKAVINDKSYKENIMRLSSLHKDRPVE
                     PLDLAVFWVEFVMRHKGAPHLRPAAHDLTWYQYHSLDVIGFLLAVVLTVAFITFKCCA
                     YGYRKCLGKKGRVKKAHKSKTH"
     TATA_signal     52972. .52988
                     /gene="UGT1A10"
     gene            52972. .>187313
                     /gene="UGT1A10"
     mRNA            join(<53103. .53957,181788. .181919,182603. .182690,
                     182974. .183193,187016. .>187313)
                     /gene="UGT1A10"
                     /product="UDP glucuronosyltransferase 1A10"
     CDS             join(53103. .53957,181788. .181919,182603. .182690,
                     182974. .183193,187016. .187313)
                     /gene="UGT1A10"
                     /codon_start=1
                     /product="UDP glucuronosyltransferase 1A10"
                     /protein_id="AAG30417.1"
                     /db_xref="GI:11118742"
                     /translation="MARAGWTSPVPLCVCLLLTCGFAEAGKLLVVPMDGSHWFTMQSV
                     VEKLILRGHEVVVVMPEVSWQLERSLNCTVKTYSTSYTLEDQNREFMVFAHAQWKAQA
                     QSIFSLLMSSSSGFLDLFFSHCRSLFNDRKLVEYLKESSFDAVFLDPFDTCGLIVAKY
                     FSLPSVVFTRGIFCHHLEEGAQCPAPLSYVPNDLLGFSDAMTFKERVWNHIVHLEDHL
                     FCQYLFRNALEIASEILQTPVTAYDLYSHTSIWLLRTDFVLDYPKPVMPNMIFIGGIN
                     CHQGKPLPMEFEAYINASGEHGIVVFSLGSMVSEIPEKKAMAIADALGKIPQTVLWRY
                     TGTRPSNLANNTILVKWLPQNDLLGHPMTRAFITHAGSHGVYESICNGVPMVMMPLFG
                     DQMDNAKRMETKGAGVTLNVLEMTSEDLENALKAVINDKSYKENIMRLSSLHKDRPVE
                     PLDLAVFWVEFVMRHKGAPHLRPAAHDLTWYQYHSLDVIGFLLAVVLTVAFITFKCCA
                     YGYRKCLGKKGRVKKAHKSKTH"
     gene            64509. .65608
                     /gene="UGT1A13p"
                     /note="UDP glucuronosyltransferase 1A13"
                     /pseudo
     CDS             join(64509. .65011,65353. .65608)
                     /gene="UGT1A13p"
                     /codon_start=1
                     /pseudo
     gene            88416. .>187313
                     /gene="UGT1A9"
     TATA_signal     88416. .88427
                     /gene="UGT1A9"
     mRNA            join(<88543. .89397,181788. .181919,182603. .182690,
                     182974. .183193,187016. .>187313)
                     /gene="UGT1A9"
                     /product="UDP glucuronosyltransferase 1A9"
     CDS             join(88543. .89397,181788. .181919,182603. .182690,
                     182974. .183193,187016. .187313)
                     /gene="UGT1A9"
                     /codon_start=1
                     /product="UDP glucuronosyltransferase 1A9"
                     /protein_id="AAG30418.1"
                     /db_xref="GI:11118743"
                     /translation="MACTGWTSPLPLCVCLLLTCGFAEAGKLLVVPMDGSHWFTMRSV
                     VEKLILRGHEVVVVMPEVSWQLGRSLNCTVKTYSTSYTLEDLDREFKAFAHAQWKAQV
                     RSIYSLLMGSYNDIFDLFFSNCRSLFKDKKLVEYLKESSFDAVFLDPFDNCGLIVAKY
                     FSLPSVVFARGILCHYLEEGAQCPAPLSYVPRILLGFSDAMTFKERVRNHIMHLEEHL
                     LCHRFFKNALEIASEILQTPVTEYDLYSHTSIWLLRTDFVLDYPKPVMPNMIFIGGIN
                     CHQGKPLPMEFEAYINASGEHGIVVFSLGSMVSEIPEKKAMAIADALGKIPQTVLWRY
                     TGTRPSNLANNTILVKWLPQNDLLGHPMTRAFITHAGSHGVYESICNGVPMVMMPLFG
                     DQMDNAKRMETKGAGVTLNVLEMTSEDLENALKAVINDKSYKENIMRLSSLHKDRPVE
                     PLDLAVFWVEFVMRHKGAPHLRPAAHDLTWYQYHSLDVIGFLLAVVLTVAFITFKCCA
                     YGYRKCLGKKGRVKKAHKSKTH"
     TATA_signal     98424. .98438
                     /gene="UGT1A7"
     gene            98424. .>187313
                     /gene="UGT1A7"
     mRNA            join(<98552. .99406,181788. .181919,182603. .182690,
                     182974. .183193,187016. .>187313)
                     /gene="UGT1A7"
                     /product="UDP glucuronosyltransferase 1A7"
     CDS             join(98552. .99406,181788. .181919,182603. .182690,
                     182974. .183193,187016. .187313)
                     /gene="UGT1A7"
                     /codon_start=1
                     /product="UDP glucuronosyltransferase 1A7"
                     /protein_id="AAG30419.1"
                     /db_xref="GI:11118744"
                     /translation="MARAGWTGLLPLYVCLLLTCGFAKAGKLLVVPMDGSHWFTMQSV
                     VEKLILRGHEVVVVMPEVSWQLGRSLNCTVKTYSTSYTLEDQDREFMVFADARWTAPL
                     RSAFSLLTSSSNGIFDLFFSNCRSLFKDKKLVEYLKESCFDAVFLDPFDACGLIVAKY
                     FSLPSVVFARGIFCHYLEEGAQCPAPLSYVPRLLLGFSDAMTFKERVRNHIMHLEEHL
                     FCPYFFKNVLEIASEILQTPVTAYDLYSHTSIWLLRTDFVLEYPKPVMPNMIFIGGIN
                     CHQGKPVPMEFEAYINASGEHGIVVFSLGSMVSEIPEKKAMAIADALGKIPQTVLWRY
                     TGTRPSNLANNTILVKWLPQNDLLGHPMTRAFITHAGSHGVYESICNGVPMVMMPLFG
                     DQMDNAKRMETKGAGVTLNVLEMTSEDLENALKAVINDKSYKENIMRLSSLHKDRPVE
                     PLDLAVFWVEFVMRHKGAPHLRPAAHDLTWYQYHSLDVIGFLLAVVLTVAFITFKCCA
                     YGYRKCLGKKGRVKKAHKSKTH"
     gene            109467. .>187313
                     /gene="UGT1A6"
     TATA_signal     109467. .109473
                     /gene="UGT1A6"
     mRNA            join(<109610. .110470,181788. .181919,182603. .182690,
                     182974. .183193,187016. .>187313)
                     /gene="UGT1A6"
                     /product="UDP glucuronosyltransferase 1A6"
     CDS             join(109610. .110470,181788. .181919,182603. .182690,
                     182974. .183193,187016. .187313)
                     /gene="UGT1A6"
                     /codon_start=1
                     /product="UDP glucuronosyltransferase 1A6"
                     /protein_id="AAG30420.1"
                     /db_xref="GI:11118745"
                     /translation="MACLLRSFQRISAGVFFLALWGMVVGDKLLVVPQDGSHWLSMKD
                     IVEVLSDRGHEIVVVVPEVNLLLKESKYYTRKIYPVPYDQEELKNRYQSFGNNHFAER
                     SPLTAPQTEYRNNMIVIGLYFINCQSLLQDRDTLNFFKESKFDALFTDPALPCGVILA
                     EYLGLPSVYLFRGFPCSLEHTFSRSPDPVSYIPRCYTKFSDHMTFSQRVANFLVNLLE
                     PYLFYCLFSKYEELASAVLKRDVDIITLYQKVSVWLLRYDFVLEYPRPVMPNMVFIGG
                     INCKKRKDLSQEFEAYINASGEHGIVVFSLGSMVSEIPEKKAMAIADALGKIPQTVLW
                     RYTGTRPSNLANNTILVKWLPQNDLLGHPMTRAFITHAGSHGVYESICNGVPMVMMPL
                     FGDQMDNAKRMETKGAGVTLNVLEMTSEDLENALKAVINDKSYKENIMRLSSLHKDRP
                     VEPLDLAVFWVEFVMRHKGAPHLRPAAHDLTWYQYHSLDVIGFLLAVVLTVAFITFKC
                     CAYGYRKCLGKKGRVKKAHKSKTH"
     gene            129525. .>187313
                     /gene="UGT1A5"
     TATA_signal     129525. .129531
                     /gene="UGT1A5"
     mRNA            join(<129600. .130466,181788. .181919,182603. .182690,
                     182974. .183193,187016. .>187313)
                     /gene="UGT1A5"
                     /product="UDP glucuronosyltransferase 1A5"
     CDS             join(129600. .130466,181788. .181919,182603. .182690,
                     182974. .183193,187016. .187313)
```

/gene="UGT1A5"
/codon_start=1
/product="UDP glucuronosyltransferase 1A5"
/protein_id="AAG30421.1"
/db_xref="GI:11118746"
/translation="MATGLQVPLPQLATGLLLLLSVQPWAESGKVLVVPTDGSHWLSM
REALRDLHARGHQVVVLTLEVNMYIKEENFFTLTTYAISWTQDEFDRLLLGHTQSFFE
TEHLLMKFSRRMAIMNNMSLIIHRSCVELLHNEALIRHLHATSFDVVLTDPFHLCAAV
LAKYLSIPAVFFLRNIPCDLDFKGTQCPNPSSYIPRLLTTNSDHMTFLQRVKNMLYPL
ALSYLCHAVSAPYASLASELFQREVSVVDLVSHASVWLFRGDFVMDYPRPIMPNMVFI
GGINCANGKPLSQEFEAYINASGEHGIVVFSLGSMVSEIPEKKAMAIADALGKIPQTV
LWRYTGTRPSNLANNTILVKWLPQNDLLGHPMTRAFITHAGSHGVYESICNGVPMVMM
PLFGDQMDNAKRMETKGAGVTLNVLEMTSEDLENALKAVINDKSYKENIMRLSSLHKD
RPVEPLDLAVFWVEFVMRHKGAPHLRPAAHDLTWYQYHSLDVIGFLLAVVLTVAFITF
KCCAYGYRKCLGKKGRVKKAHKSKTH"

Query Match 100.0%; Score 21; DB 89; Length 198872;
Best Local Similarity 100.0%; Pred. No. 0.83;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ccactgggatcaacagtatct 21
||||||||||||||||||||
Db 175148 CCACTGGGATCAACAGTATCT 175128

RESULT 12
AY029169/c
LOCUS AY029169 1602 bp mRNA PRI 16-APR-2001
DEFINITION Macaca mulatta UDP-glucuronosyltransferase UGT1A01 mRNA, complete cds.
ACCESSION AY029169
VERSION AY029169.1 GI:13641264
KEYWORDS .
SOURCE rhesus monkey.
ORGANISM Macaca mulatta
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae; Cercopithecinae; Macaca.
REFERENCE 1 (bases 1 to 1602)
AUTHORS Dean,B.J., Zhao,S. and King,C.
TITLE Nucleotide and Deduced Amino Acid Sequence of a Novel UDP-Glucuronosyltransferase 1A Isoform Cloned from Male Rhesus Monkey Liver
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1602)
AUTHORS Dean,B.J., Zhao,S. and King,C.
TITLE Direct Submission
JOURNAL Submitted (02-APR-2001) Drug Metabolism, Merck & Co. Inc., Mail Drop RY80L-109, PO Box 2000, Rahway, NJ 07065, USA
FEATURES Location/Qualifiers
source 1. .1602
/organism="Macaca mulatta"
/db_xref="taxon:9544"
/sex="male"
CDS 1. .1602
/codon_start=1
/product="UDP-glucuronosyltransferase UGT1A01"
/protein_id="AAK31597.1"
/db_xref="GI:13641265"
/translation="MAVESQGRHPLVLGLLLCVLGPVLCHAGKMLLIPVDGSHWLSML
GTIQQLQQRGHEIVVLAPDASLYIREGAFYTLKTYPVPFQREDVKESFVSLGHNVFEN
DSFLQRVIKTYKKIKKDSAMLLSGCSHLLHNKELMASLAESSFDVMLTDPFLPCGPIV
AQYLSLPTVFFLNALPCSLESEATQCPNPFSYVPRPLSAHSDHMTFLQRVKSMLIAFS
QNFLCDVVYSPYATLASEFLQREVTVQNLLSSASVWLLRSDFVKDYPRPIMPNMAFIG
GINCLHQSPLSQEFEAYINASGEHGIVVFSLGSMVAEIPEKKAMAIADALGKIPQTVL
WRYTGTPPSNLANNTILVKWLPQNDLLGHPMTRAFITHAGSHGIYEGICNGVPMVMMP
LFGDQMDNAKRMETKGAGVTLNVLEMTSEDLENALKAVINDKSYKENIMHLSSLHKDR
PVEPLDLAVFWVEFVMRHKGAPHLRPAAHDLTWYQYHSLDVIGFLLAIVLTVAFIAFK
CCAYGYRKCFGKKGRVKKAHKSKTH"
BASE COUNT 383 a 417 c 390 g 412 t
ORIGIN

Query Match 92.4%; Score 19.4; DB 91; Length 1602;
Best Local Similarity 95.2%; Pred. No. 6;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ccactgggatcaacagtatct 21
||||||||||||||||| ||||
Db 106 CCACTGGGATCAACAGCATCT 86

RESULT 13
AF104339/c
LOCUS AF104339 1647 bp mRNA PRI 13-DEC-1999
DEFINITION Macaca fascicularis UDP-glucuronosyltransferase UGT1A01 mRNA, complete cds.
ACCESSION AF104339
VERSION AF104339.1 GI:6537143
KEYWORDS .
SOURCE crab-eating macaque.
ORGANISM Macaca fascicularis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae; Cercopithecinae; Macaca.
REFERENCE 1 (bases 1 to 1647)
AUTHORS Albert,C., Vallee,M., Beaudry,G., Belanger,A. and Hum,D.W.
TITLE Direct Submission
JOURNAL Submitted (05-NOV-1998) Molecular Endocrinology, Laval University, 2705 Laurier Boulevard, Sainte-Foy, Que G1V4G2, Canada
FEATURES Location/Qualifiers
source 1. .1647
/organism="Macaca fascicularis"
/db_xref="taxon:9541"
CDS 38. .1639
/codon_start=1
/product="UDP-glucuronosyltransferase UGT1A01"
/protein_id="AAF15549.1"
/db_xref="GI:6537144"
/translation="MAVESQGRHPLVLGLLLCVLGPVLCHAGKMLLIPVDGSHWLSML
GTIQQLQQRGHEIVVLAPDASLYIREGAFYTLKTYPVPFQREDVKESFVSLGHNVFEN
DSFLRRVIKTYKKIKKDSAMLLSGCSHLLHNKELMASLAESSFDVMLTDPFLPCGPIV
AQYLSLPTVFFLNALPCSLESEATQCPNPFSYVPRPLSAHSDHMTFLQRVKNMLIAFS
QNFLCDVVYSPYATLASEFLQREVTVQNLLSSASVWLLRSDFVKDYPRPIMPNMAFIG
GINCLHQSPLSQEFEAYINASGEHGIVVFSLGSMVAEIPEKKAMAIADALGKIPQTVL
WRYTGTPPSNLANNTILVKWLPQNDLLGHPMTRAFITHAGSHGIYEGICNGVPMVMMP
LFGDQMDNAKRMETKGAGVTLNVLEMTSEDLENALKAVINDKSYKENIMHLSSLHKDR
PVEPLDLAVFWVEFVMRHKGAPHLRPAAHDLTWYQYHSLDVIGFLLAIVLTVAFIAFK
CCAYGYRKCFGKKGRVKKAHKSKTH"
BASE COUNT 393 a 431 c 409 g 414 t
ORIGIN

Query Match 92.4%; Score 19.4; DB 88; Length 1647;
Best Local Similarity 95.2%; Pred. No. 6;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ccactgggatcaacagtatct 21
||||||||||||||||| ||||
Db 143 CCACTGGGATCAACAGCATCT 123

RESULT 14
AC007905
LOCUS AC007905 104480 bp DNA HTG 03-OCT-2000
DEFINITION Homo sapiens chromosome 16q24.3 clone RP4-754F23, WORKING DRAFT SEQUENCE, 35 ordered pieces.
ACCESSION AC007905
VERSION AC007905.2 GI:10280722
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 104480)
AUTHORS Kremmidiotis,G., Gardner,A.E., Callen,D.F. and Sutherland,G.R.

```
  TITLE     Large Scale Sequencing of the Chromosome 16 region q24.3
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 104480)
  AUTHORS   Kremmidiotis,G., Gardner,A.E., Callen,D.F. and Sutherland,G.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-JUN-1999) Cytogenetics & Molecular Genetics, Women's
            & Children's Hospital, 72 King William Rd, Adelaide, SA 5006,
            Australia
COMMENT     On Sep 23, 2000 this sequence version replaced gi:5174819.
            ------------------------------------------------------------------
            Genome Centre : Cytogenetics & Molecular Genetics
            Centre Code : CMGWCH
            Website:http://www.wch.sa.gov.au/labmedic/genetics/sequencing.html
            ------------------------------------------------------------------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 35 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *        1     1479: contig of 1479 bp in length
            *     1480     1487: gap of unknown length
            *     1488     6699: contig of 5212 bp in length
            *     6700     6707: gap of unknown length
            *     6708    10905: contig of 4198 bp in length
            *    10906    10912: gap of unknown length
            *    10913    11733: contig of 821 bp in length
            *    11734    11740: gap of unknown length
            *    11741    12602: contig of 862 bp in length
            *    12603    12609: gap of unknown length
            *    12610    13856: contig of 1247 bp in length
            *    13857    13863: gap of unknown length
            *    13864    20018: contig of 6155 bp in length
            *    20019    20025: gap of unknown length
            *    20026    21942: contig of 1917 bp in length
            *    21943    21949: gap of unknown length
            *    21950    23920: contig of 1971 bp in length
            *    23921    23927: gap of unknown length
            *    23928    24860: contig of 933 bp in length
            *    24861    24867: gap of unknown length
            *    24868    29279: contig of 4412 bp in length
            *    29280    29286: gap of unknown length
            *    29287    29772: contig of 486 bp in length
            *    29773    29779: gap of unknown length
            *    29780    32411: contig of 2632 bp in length
            *    32412    32418: gap of unknown length
            *    32419    33145: contig of 727 bp in length
            *    33146    33152: gap of unknown length
            *    33153    34862: contig of 1710 bp in length
            *    34863    34869: gap of unknown length
            *    34870    36643: contig of 1774 bp in length
            *    36644    36650: gap of unknown length
            *    36651    37348: contig of 698 bp in length
            *    37349    37355: gap of unknown length
            *    37356    42174: contig of 4819 bp in length
            *    42175    42181: gap of unknown length
            *    42182    45604: contig of 3423 bp in length
            *    45605    45611: gap of unknown length
            *    45612    48485: contig of 2874 bp in length
            *    48486    48492: gap of unknown length
            *    48493    49845: contig of 1353 bp in length
            *    49846    49852: gap of unknown length
            *    49853    51199: contig of 1347 bp in length
            *    51200    51206: gap of unknown length
            *    51207    51513: contig of 307 bp in length
            *    51514    51520: gap of unknown length
            *    51521    52180: contig of 660 bp in length
            *    52181    52187: gap of unknown length
            *    52188    60026: contig of 7839 bp in length
            *    60027    60033: gap of unknown length
            *    60034    64699: contig of 4666 bp in length
            *    64700    64706: gap of unknown length
            *    64707    69136: contig of 4430 bp in length
            *    69137    69143: gap of unknown length
            *    69144    73347: contig of 4204 bp in length
            *    73348    73354: gap of unknown length
            *    73355    82462: contig of 9108 bp in length
            *    82463    82469: gap of unknown length
            *    82470    84166: contig of 1697 bp in length
            *    84167    84173: gap of unknown length
            *    84174    84822: contig of 649 bp in length
            *    84823    84829: gap of unknown length
            *    84830    99511: contig of 14682 bp in length
            *    99512    99518: gap of unknown length
            *    99519   101087: contig of 1569 bp in length
            *   101088   101094: gap of unknown length
            *   101095   103316: contig of 2222 bp in length
            *   103317   103323: gap of unknown length
            *   103324   104480: contig of 1157 bp in length.
FEATURES             Location/Qualifiers
     source          1. .104480
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="16q24.3"
                     /clone="RP4-754F23"
BASE COUNT    23140 a  28756 c  29631 g  22706 t    247 others
ORIGIN


  Query Match           84.8%;  Score 17.8;  DB 60;  Length 104480;
  Best Local Similarity  90.5%;  Pred. No. 52;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 ccactgggatcaacagtatct 21
          ||||||||||||||| ||| |
Db  20902 CCACTGGGATCAACATTATTT 20922


RESULT  15
HSDJ137K2/c
LOCUS       HSDJ137K2  187507 bp    DNA              PRI       24-JUL-2000
DEFINITION  Human DNA sequence from clone RP1-137K2 on chromosome 6q25.1-25.3.
            Contains part of the gene for a novel protein similar to B120
            (contains KIAA1235), ESTs, STSs, GSSs and two putative CpG islands,
            complete sequence.
ACCESSION   AL049820
VERSION     AL049820.23  GI:8247261
KEYWORDS    HTG; B120; CpG island; KIAA1235.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 187507)
  AUTHORS   Sycamore,N.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-JUL-2000) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Jun 4, 2000 this sequence version replaced gi:8018160.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence has been finished according to sequence map criteria
            as follows. An attempt is made to resolve all sequencing problems,
            such as compressions and repeats, but not necessarily within known
            annotated human repeat sequence elements (e.g. Alu). Where the
            sequence is ambiguous, there is an annotation using the 'unsure'
            feature key.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
```

Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping Group. Further information can be found at http://www.sanger.ac.uk/HGP/Chr6
RP1-137K2 is from the library RPCI-1 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see http://bacpac.med.buffalo.edu/
VECTOR: pCYPAC2
This sequence is the entire insert of clone RP1-137K2 The true right end of clone RP1-80E10 is at 27741 in this sequence.

FEATURES Location/Qualifiers

source 1. .187507
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/map="q25.1-25.3"
/clone="RP1-137K2"
/clone_lib="RPCI-1"

repeat_region 1. .230
/note="AluSq repeat: matches 2. .237 of consensus"

misc_feature complement(1808. .2197)
/note="match: GSS: Em:AQ106481"

repeat_region 2678. .2782
/note="MER53 repeat: matches 85. .189 of consensus"

misc_feature complement(3848. .4325)
/note="match: GSS: Em:B71907"

repeat_region 5108. .5417
/note="AluY repeat: matches 1. .297 of consensus"

repeat_region 6157. .6278
/note="MIR repeat: matches 15. .153 of consensus"

repeat_region 6589. .6661
/note="L2 repeat: matches 2671. .2741 of consensus"

repeat_region 6663. .6727
/note="AluJ/FLAM repeat: matches 1. .66 of consensus"

repeat_region 6768. .7057
/note="L2 repeat: matches 2356. .2661 of consensus"

misc_feature 7142. .7944
/note="match: GSS: Em:AQ747336"

repeat_region 7589. .7889
/note="AluSx repeat: matches 1. .305 of consensus"

repeat_region 9589. .9699
/note="L2 repeat: matches 2579. .2708 of consensus"

repeat_region 9826. .10165
/note="L2 repeat: matches 2080. .2417 of consensus"

repeat_region 10580. .10752
/note="AluJo repeat: matches 132. .302 of consensus"

repeat_region 10795. .10905
/note="AluJb repeat: matches 48. .158 of consensus"

repeat_region 10975. .11081
/note="AluJo repeat: matches 14. .125 of consensus"

repeat_region 11763. .11796
/note="17 copies 2 mer tt 82% conserved"

repeat_region 11797. .12023
/note="AluSg/x repeat: matches 86. .312 of consensus"

repeat_region 12067. .12348
/note="AluSg repeat: matches 1. .280 of consensus"

misc_feature 14350. .14877
/note="match: GSS: Em:AQ197120"

misc_feature 14357. .14684
/note="match: GSS: Em:AQ312897"

repeat_region 14623. .14676
/note="Alu repeat: matches 249. .302 of consensus"

repeat_region 14677. .14870
/note="AluY repeat: matches 1. .302 of consensus"

repeat_region 15034. .15291
/note="L1MB3 repeat: matches 5921. .6181 of consensus"

repeat_region 16601. .16797
/note="MIR repeat: matches 14. .219 of consensus"

repeat_region 17172. .17231
/note="30 copies 2 mer tt 81% conserved"

repeat_region 17174. .17225
/note="4 copies 13 mer 84% conserved"

repeat_region 17234. .17514
/note="AluSc repeat: matches 1. .283 of consensus"

repeat_region 17515. .17564
/note="25 copies 2 mer tt 80% conserved"

repeat_region 17517. .17556
/note="8 copies 5 mer ttgtt 87% conserved"

repeat_region 18805. .18879
/note="L1M4c repeat: matches 1685. .1759 of consensus"

repeat_region 19556. .19625
/note="MIR repeat: matches 78. .146 of consensus"

repeat_region 19765. .20073
/note="AluSp repeat: matches 1. .310 of consensus"

misc_feature complement(20235. .20678)
/note="match: GSS: Em:AQ465222"

repeat_region 22127. .22450
/note="L2 repeat: matches 1579. .1915 of consensus"

repeat_region 23241. .23349
/note="MER33 repeat: matches 34. .142 of consensus"

repeat_region 23350. .23846
/note="MER6 repeat: matches 1. .865 of consensus"

repeat_region 23847. .24019
/note="MER33 repeat: matches 141. .324 of consensus"

misc_feature 24447. .24872
/note="match: GSS: Em:AQ526363"

misc_feature 24458. .24762
/note="match: GSS: Em:AQ665701"

repeat_region 24739. .25035
/note="AluJo repeat: matches 1. .291 of consensus"

repeat_region 25108. .25407
/note="AluSg repeat: matches 2. .302 of consensus"

repeat_region 25938. .26203
/note="L2 repeat: matches 2002. .2292 of consensus"

repeat_region 26263. .26352
/note="MIR repeat: matches 91. .192 of consensus"

misc_feature 27822. .29153
/note="CpG island"
/evidence=not_experimental

misc_feature 27853. .28381
/note="match: GSS: Em:AQ281427"

misc_feature complement(28203. .28600)
/note="match: GSS: Em:AQ937194"

misc_feature 28659. .29093
/note="match: GSS: Em:AQ937195"

repeat_region 29028. .29118
/note="7 copies 13 mer 70% conserved"

misc_feature 29721. .30213
/note="match: GSS: Em:AQ214419"

misc_feature 29780. .30137
/note="match: GSS: Em:AQ099774"

misc_feature 29812. .30102
/note="match: GSS: Em:AQ096409"

repeat_region 31375. .31688
/note="AluSx repeat: matches 1. .308 of consensus"

repeat_region 32258. .32813
/note="L1ME1 repeat: matches 5549. .6053 of consensus"

repeat_region 32823. .33107
/note="AluSx repeat: matches 1. .297 of consensus"

repeat_region 33581. .33709
/note="FLAM_A repeat: matches 1. .128 of consensus"

misc_feature complement(33668. .34018)
/note="match: STS: Em:G24299"

repeat_region 34949. .34972
/note="12 copies 2 mer tg 95% conserved"

repeat_region 36306. .36476
/note="L2 repeat: matches 2571. .2750 of consensus"

repeat_region 36527. .36791
/note="L2 repeat: matches 2159. .2428 of consensus"

repeat_region 36931. .37354
/note="MLT1H repeat: matches 57. .504 of consensus"

repeat_region 38691. .38991
/note="AluSx repeat: matches 1. .296 of consensus"

```
     repeat_region   44547. .44716
                     /note="L2 repeat: matches 2009. .2202 of consensus"
     repeat_region   44907. .45038
                     /note="2 copies 66 mer 96% conserved"
     repeat_region   45805. .46069
                     /note="AluJb repeat: matches 1. .294 of consensus"
     repeat_region   46214. .46442
                     /note="AluYb8 repeat: matches 83. .311 of consensus"
     repeat_region   46533. .46827
                     /note="5 copies 59 mer 70% conserved"
     repeat_region   46550. .46747
                     /note="3 copies 66 mer 74% conserved"
     repeat_region   46574. .46825
                     /note="6 copies 42 mer 66% conserved"
     repeat_region   46584. .46830
                     /note="19 copies 13 mer 61% conserved"
     repeat_region   46714. .46771
                     /note="29 copies 2 mer tc 87% conserved"
     repeat_region   46835. .47132
                     /note="AluSc repeat: matches 1. .308 of consensus"
     misc_feature    47007. .47546

  Query Match             84.8%;  Score 17.8;  DB 93;  Length 187507;
  Best Local Similarity   90.5%;  Pred. No. 53;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 ccactgggatcaacagtatct 21
           ||||||| |||| ||||||||
Db 118886 CCACTGGCATCAGCAGTATCT 118866


Search completed: July 25, 2001, 05:16:58
Job time: 9221 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: July 25, 2001, 02:43:17 ; Search time 1290.33 Seconds
(without alignments)
251.736 Million cell updates/sec

Title: US-09-142-095-1
Perfect score: 21
Sequence: 1 aagtgaactccctgctacctt 21

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues

Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_ba3:*
4: gb_in1:*
5: gb_in2:*
6: gb_in3:*
7: gb_om:*
8: gb_ov:*
9: gb_pat1:*
10: gb_pat2:*
11: gb_ph:*
12: gb_pl1:*
13: gb_pl2:*
14: gb_pl3:*
15: gb_pl4:*
16: em_ba1:*
17: em_ba2:*
18: em_fun:*
19: em_htgo_hum:*
20: em_htgo_inv:*
21: em_htgo_rod:*
22: em_htg_hum1:*
23: em_htg_hum2:*
24: em_htg_hum3:*
25: em_htg_hum4:*
26: em_htg_hum5:*
27: em_htg_hum6:*
28: em_htg_hum7:*
29: em_htg_hum8:*
30: em_htg_inv1:*
31: em_htg_inv2:*
32: em_htg_other:*
33: em_htg_rod:*
34: em_hum1:*
35: em_hum2:*
36: em_hum3:*
37: em_hum4:*
38: em_hum5:*
39: em_hum6:*
40: em_hum7:*
41: em_in:*
42: em_om:*
43: em_or:*
44: em_ov:*
45: em_pat:*
46: em_ph:*
47: em_pl:*
48: em_ro:*
49: em_sts:*
50: em_sy:*
51: em_un:*
52: em_vi:*
53: gb_sts1:*
54: gb_sts2:*
55: gb_sts3:*
56: gb_sy:*
57: gb_un:*
58: gb_vi1:*
59: gb_vi2:*
60: gb_htg1:*
61: gb_htg2:*
62: gb_htg3:*
63: gb_htg4:*
64: gb_htg5:*
65: gb_htg6:*
66: gb_htg7:*
67: gb_htg8:*
68: gb_htg9:*
69: gb_htg10:*
70: gb_htg11:*
71: gb_htg12:*
72: gb_htg13:*
73: gb_htg14:*
74: gb_htg15:*
75: gb_htg16:*
76: gb_htg17:*
77: gb_htg18:*
78: gb_htg19:*
79: gb_htg20:*
80: gb_htg21:*
81: gb_htg22:*
82: gb_htg23:*
83: gb_htg24:*
84: gb_htg25:*
85: gb_pr1:*
86: gb_pr2:*
87: gb_pr3:*
88: gb_pr4:*
89: gb_pr5:*
90: gb_pr6:*
91: gb_pr7:*
92: gb_pr8:*
93: gb_pr9:*
94: gb_ro1:*
95: gb_ro2:*
96: gb_in4:*
97: gb_pr10:*
98: em_ba3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	21	100.0	21	9	A65500	A65500 Sequence 1
2	21	100.0	208	89	AF135471	AF135471 Saimiri b
3	21	100.0	234	89	AF135462	AF135462 Pan panis
4	21	100.0	234	89	AF135463	AF135463 Pan trogl
5	21	100.0	531	89	AF352795	AF352795 Homo sapi
6	21	100.0	541	89	AF180372	AF180372 Homo sapi
7	21	100.0	620	9	A65504	A65504 Sequence 5
8	21	100.0	3341	91	D87674	D87674 Homo sapien

c	9	21	100.0	68770	86	AC006985	AC006985 Homo sapi
c	10	21	100.0	176619	70	AC026497	AC026497 Homo sapi
	11	21	100.0	198872	89	AF297093	AF297093 Homo sapi
c	12	18.4	87.6	136868	86	AC008278	AC008278 Homo sapi
	13	17.8	84.8	7218	91	AP000293	AP000293 Homo sapi
	14	17.8	84.8	46831	77	AC087204	AC087204 Homo sapi
c	15	17.8	84.8	47972	90	AL391843	AL391843 Human DNA
	16	17.8	84.8	57000	88	AC069278	AC069278 Homo sapi
	17	17.8	84.8	70068	77	AC090246	AC090246 Homo sapi
	18	17.8	84.8	100000	91	AP000043	AP000043 Homo sapi
	19	17.8	84.8	100000	91	AP000111	AP000111 Homo sapi
	20	17.8	84.8	100000	91	AP000187	AP000187 Homo sapi
	21	17.8	84.8	105733	75	AC074136	AC074136 Homo sapi
	22	17.8	84.8	125428	88	AC073898	AC073898 Homo sapi
	23	17.8	84.8	139493	65	AC018537	AC018537 Homo sapi
	24	17.8	84.8	145540	64	AC015962	AC015962 Homo sapi
c	25	17.8	84.8	145540	64	AC015962	AC015962 Homo sapi
c	26	17.8	84.8	153234	70	AC026947	AC026947 Homo sapi
	27	17.8	84.8	156486	62	AC011657	AC011657 Homo sapi
c	28	17.8	84.8	156771	82	AL590223	AL590223 Homo sapi
c	29	17.8	84.8	165987	66	AC021201	AC021201 Homo sapi
	30	17.8	84.8	167195	87	AC009044	AC009044 Homo sapi
c	31	17.8	84.8	170916	76	AC079414	AC079414 Homo sapi
	32	17.8	84.8	173985	76	AC084085	AC084085 Homo sapi
	33	17.8	84.8	174788	82	AP000756	AP000756 Homo sapi
c	34	17.8	84.8	175393	73	AC068412	AC068412 Homo sapi
	35	17.8	84.8	175587	77	AC087616	AC087616 Homo sapi
	36	17.8	84.8	176107	83	AP002393	AP002393 Homo sapi
	37	17.8	84.8	176317	62	AC011468	AC011468 Homo sapi
	38	17.8	84.8	182815	62	AC011386	AC011386 Homo sapi
c	39	17.8	84.8	183279	75	AC079014	AC079014 Homo sapi
	40	17.8	84.8	191602	73	AC068920	AC068920 Homo sapi
	41	17.8	84.8	340000	91	AP001716	AP001716 Homo sapi
c	42	17.4	82.9	1073	53	CNS06R3X	AL411475 T7 end of
	43	17	81.0	29605	12	AB020747	AB020747 Arabidops
	44	17	81.0	105223	12	AC007399	AC007399 Arabidops
	45	17	81.0	156970	74	AC073330	AC073330 Homo sapi

ALIGNMENTS

```
RESULT   1
A65500
LOCUS       A65500          21 bp    DNA             PAT       29-MAR-1999
DEFINITION  Sequence 1 from Patent WO9732042.
ACCESSION   A65500
VERSION     A65500.1  GI:4531235
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 21)
  AUTHORS   Burchell,B.
  TITLE     DRUG TRIAL ASSAY SYSTEM
  JOURNAL   Patent: WO 9732042-A 1 04-SEP-1997;
            UNIV DUNDEE (GB)
COMMENT     Other publication AU 2224197 19970916.
FEATURES             Location/Qualifiers
     source          1. .21
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT        5 a      7 c      3 g      6 t
ORIGIN


  Query Match           100.0%;  Score 21;  DB 9;  Length 21;
  Best Local Similarity  100.0%;  Pred. No. 0.96;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 aagtgaactccctgctacctt 21
          |||||||||||||||||||||
Db      1 AAGTGAACTCCCTGCTACCTT 21
```

```
RESULT   2
AF135471
LOCUS       AF135471        208 bp    DNA             PRI       21-NOV-1999
DEFINITION  Saimiri boliviensis UDP-glucuronosyltransferase 1A1 (UGT1A1) gene,
            promoter region and partial cds.
ACCESSION   AF135471
VERSION     AF135471.1  GI:6456559
KEYWORDS    .
SOURCE      Bolivian squirrel monkey.
  ORGANISM  Saimiri boliviensis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Platyrrhini; Cebidae; Cebinae;
            Saimiri.
REFERENCE   1  (bases 1 to 208)
  AUTHORS   Hall,D., Ybazeta,G., Destro-Bisol,G., Petzl-Erler,M.L. and Di
            Rienzo,A.
  TITLE     Variability at the uridine diphosphate glucuronosyltransferase 1A1
            promoter in human populations and primates
  JOURNAL   Pharmacogenetics (1999) In press
REFERENCE   2  (bases 1 to 208)
  AUTHORS   Ybazeta,G., Hall,D. and Di Rienzo,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-MAR-1999) Human Genetics, University of Chicago, 924
            E. 57th Street, Chicago, IL 60637, USA
FEATURES             Location/Qualifiers
     source          1. .208
                     /organism="Saimiri boliviensis"
                     /db_xref="taxon:27679"
     mRNA            <142. .>208
                     /gene="UGT1A1"
                     /product="UDP-glucuronosyltransferase 1A1"
     gene            <142. .>208
                     /gene="UGT1A1"
     CDS             142. .>208
                     /gene="UGT1A1"
                     /codon_start=1
                     /product="UDP-glucuronosyltransferase 1A1"
                     /protein_id="AAF09182.1"
                     /db_xref="GI:6456560"
                     /translation="MAAESRGGCQLVLGLLLCVLGP"
BASE COUNT       44 a     52 c     58 g     54 t
ORIGIN


  Query Match           100.0%;  Score 21;  DB 89;  Length 208;
  Best Local Similarity  100.0%;  Pred. No. 0.85;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 aagtgaactccctgctacctt 21
          |||||||||||||||||||||
Db      3 AAGTGAACTCCCTGCTACCTT 23
```

```
RESULT   3
AF135462
LOCUS       AF135462        234 bp    DNA             PRI       21-NOV-1999
DEFINITION  Pan paniscus UDP-glucuronosyltransferase 1A1 (UGT1A1) gene,
            promoter region and partial cds.
ACCESSION   AF135462
VERSION     AF135462.1  GI:6456541
KEYWORDS    .
SOURCE      pygmy chimpanzee.
  ORGANISM  Pan paniscus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1  (bases 1 to 234)
  AUTHORS   Hall,D., Ybazeta,G., Destro-Bisol,G., Petzl-Erler,M.L. and Di
            Rienzo,A.
  TITLE     Variability at the uridine diphosphate glucuronosyltransferase 1A1
            promoter in human populations and primates
  JOURNAL   Pharmacogenetics (1999) In press
```

```
REFERENCE   2  (bases 1 to 234)
  AUTHORS   Ybazeta,G., Hall,D. and Di Rienzo,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-MAR-1999) Human Genetics, University of Chicago, 924
            E. 57th Street, Chicago, IL 60637, USA
FEATURES             Location/Qualifiers
     source          1. .234
                     /organism="Pan paniscus"
                     /db_xref="taxon:9597"
     mRNA            <148. .>234
                     /gene="UGT1A1"
                     /product="UDP-glucuronosyltransferase 1A1"
     gene            <148. .>234
                     /gene="UGT1A1"
     CDS             148. .>234
                     /gene="UGT1A1"
                     /codon_start=1
                     /product="UDP-glucuronosyltransferase 1A1"
                     /protein_id="AAF09173.1"
                     /db_xref="GI:6456542"
                     /translation="MAVESQGGRPLVLGLLLCVLGPVVSHAGK"
BASE COUNT       50 a     53 c     69 g     62 t
ORIGIN


  Query Match           100.0%;  Score 21;  DB 89;  Length 234;
  Best Local Similarity  100.0%;  Pred. No. 0.84;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 aagtgaactccctgctacctt 21
          |||||||||||||||||||||
Db      3 AAGTGAACTCCCTGCTACCTT 23


RESULT   4
AF135463
LOCUS       AF135463      234 bp    DNA             PRI       21-NOV-1999
DEFINITION  Pan troglodytes UDP-glucuronosyltransferase 1A1 (UGT1A1) gene,
            promoter region and partial cds.
ACCESSION   AF135463
VERSION     AF135463.1  GI:6456543
KEYWORDS    .
SOURCE      chimpanzee.
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1  (bases 1 to 234)
  AUTHORS   Hall,D., Ybazeta,G., Destro-Bisol,G., Petzl-Erler,M.L. and Di
            Rienzo,A.
  TITLE     Variability at the uridine diphosphate glucuronosyltransferase 1A1
            promoter in human populations and primates
  JOURNAL   Pharmacogenetics (1999) In press
REFERENCE   2  (bases 1 to 234)
  AUTHORS   Ybazeta,G., Hall,D. and Di Rienzo,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-MAR-1999) Human Genetics, University of Chicago, 924
            E. 57th Street, Chicago, IL 60637, USA
FEATURES             Location/Qualifiers
     source          1. .234
                     /organism="Pan troglodytes"
                     /db_xref="taxon:9598"
     mRNA            <148. .>234
                     /gene="UGT1A1"
                     /product="UDP-glucuronosyltransferase 1A1"
     gene            <148. .>234
                     /gene="UGT1A1"
     CDS             148. .>234
                     /gene="UGT1A1"
                     /codon_start=1
                     /product="UDP-glucuronosyltransferase 1A1"
                     /protein_id="AAF09174.1"
                     /db_xref="GI:6456544"
                     /translation="MAVESQGGRPLVLGLLLCVLGPVVSHAGK"
BASE COUNT       50 a     53 c     69 g     62 t
ORIGIN


  Query Match           100.0%;  Score 21;  DB 89;  Length 234;
  Best Local Similarity  100.0%;  Pred. No. 0.84;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 aagtgaactccctgctacctt 21
          |||||||||||||||||||||
Db      3 AAGTGAACTCCCTGCTACCTT 23


RESULT   5
AF352795
LOCUS       AF352795      531 bp    DNA             PRI       10-APR-2001
DEFINITION  Homo sapiens bilirubin UDP-glucuronosyltrasferase 1-1 (UGT1A1) gene,
            UGT1A1*1 allele, partial cds.
ACCESSION   AF352795
VERSION     AF352795.1  GI:13569708
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 531)
  AUTHORS   McKie,K., Kutlar,F., Glendenning,M. and Kutlar,A.
  TITLE     7(TA) repeat polymorphism of the TATA box of human bilirubin
            UDP-glucuronosyltransferase 1-1(UGT1A1*1) gene in a patient with
            sickle cell anemia + high bilirubinemia
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 531)
  AUTHORS   McKie,K., Kutlar,F., Glendenning,M. and Kutlar,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-FEB-2001) Medicine/Hemoglobin DNA Laboratory, Medical
            College of Georgia, 15th St., AC-1000, Augusta, GA 30912, USA
FEATURES             Location/Qualifiers
     source          1. .531
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /map="2q37"
                     /cell_type="white blood cell"
                     /tissue_type="blood"
     gene            <1. .>531
                     /gene="UGT1A1"
                     /note="UGT1"
                     /allele="UGT1A1*1"
     promoter        <1. .313
                     /gene="UGT1A1"
     TATA_signal     279. .295
                     /gene="UGT1A1"
     repeat_region   279. .>292
                     /note="repeat polymorphism compared to UGT1A1 sequence
                     presented in GenBank Accession Number AF180372; contains 7
                     ta repeats frequently detected in patients with high
                     bilirubinemia"
                     /rpt_type=tandem
                     /rpt_unit=ta
     mRNA            314. .>531
                     /gene="UGT1A1"
                     /product="bilirubin UDP-glucronosyltrasferase 1-1"
     CDS             334. .>531
                     /gene="UGT1A1"
                     /codon_start=1
                     /product="bilirubin UDP-glucronosyltrasferase 1-1"
                     /protein_id="AAK31204.1"
                     /db_xref="GI:13569709"
                     /translation="MAVESQGGRPLVLGLLLCVLGPVVSHAGKILLIPVDGSHWLSML
                     GAIQQLQQRGHEIVVLAPDASL"
BASE COUNT      120 a    121 c    137 g    153 t
ORIGIN
```

```
  Query Match           100.0%;  Score 21;  DB 89;  Length 531;
  Best Local Similarity  100.0%;  Pred. No. 0.81;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 aagtgaactccctgctacctt 21
          |||||||||||||||||||||
Db    185 AAGTGAACTCCCTGCTACCTT 205


RESULT   6
AF180372
LOCUS       AF180372      541 bp    DNA             PRI       05-OCT-1999
DEFINITION  Homo sapiens bilirubin UDP-glucuronosyltransferase 1-1 (UGT1) gene,
            UGT1*1 allele, partial cds.
ACCESSION   AF180372
VERSION     AF180372.1  GI:6010649
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 541)
  AUTHORS   Kutlar,F., Sromek,E., Leithner,C., Nechtman,J. and Kutlar,A.
  TITLE     Detection of the TATA box polymorphism of the human bilirubin
            UDP-glucuronosyltransferase 1-1 gene (UGT1*1) in a patient with
            sickle cell anemia
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 541)
  AUTHORS   Kutlar,F., Sromek,E., Leithner,C., Nechtman,J. and Kutlar,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-AUG-1999) Medicine, Hematology/Oncology-Sickle Cell
            Center, Medical College of Georgia, 15th Street, AC-1000, Augusta,
            GA 30912, USA
FEATURES             Location/Qualifiers
     source          1. .541
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /map="2q37"
                     /cell_type="white blood cells"
                     /tissue_type="whole blood"
     gene            1. .>541
                     /gene="UGT1"
                     /note="GNT1; UGT1A1"
                     /allele="UGT1*1"
     promoter        1. .313
                     /gene="UGT1"
     variation       226
                     /gene="UGT1"
                     /replace="t"
     repeat_region   279. .288
                     /note="polymorphic region"
                     /rpt_type=tandem
                     /rpt_unit-ta
     TATA_signal     279. .291
                     /gene="UGT1"
                     /note="Ritter,J.K., et al., 1992, J. Biol. Chem.,
                     267:3257-3261"
     mRNA            314. .>541
                     /gene="UGT1"
                     /product="bilirubin UDP-glucuronosyltransferase 1-1"
     CDS             330. .>541
                     /gene="UGT1"
                     /note="UDP glycosyltransferase 1"
                     /codon_start=1
                     /product="bilirubin UDP-glucuronosyltransferase 1-1"
                     /protein_id="AAF01205.1"
                     /db_xref="GI:6010650"
                     /translation="MAVESQGGRPLVLGLLLCVLGPVVSHAGKILLIPVDGSHWLSML
                     GAIQQLQQRGHEIVVLAPDASLYIRDG"
BASE COUNT      123 a    124 c    141 g    153 t
ORIGIN
```

```
  Query Match           100.0%;  Score 21;  DB 89;  Length 541;
  Best Local Similarity  100.0%;  Pred. No. 0.81;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 aagtgaactccctgctacctt 21
          |||||||||||||||||||||
Db    185 AAGTGAACTCCCTGCTACCTT 205


RESULT   7
A65504
LOCUS       A65504        620 bp    DNA             PAT       29-MAR-1999
DEFINITION  Sequence 5 from Patent WO9732042.
ACCESSION   A65504
VERSION     A65504.1  GI:4531239
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 620)
  AUTHORS   Burchell,B.
  TITLE     DRUG TRIAL ASSAY SYSTEM
  JOURNAL   Patent: WO 9732042-A 5 04-SEP-1997;
            UNIV DUNDEE (GB)
COMMENT     Other publication AU 2224197 19970916.
FEATURES             Location/Qualifiers
     source          1. .620
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      157 a    127 c    151 g    185 t
ORIGIN


  Query Match           100.0%;  Score 21;  DB 9;  Length 620;
  Best Local Similarity  100.0%;  Pred. No. 0.8;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 aagtgaactccctgctacctt 21
          |||||||||||||||||||||
Db    465 AAGTGAACTCCCTGCTACCTT 485


RESULT   8
D87674
LOCUS       D87674       3341 bp    DNA             PRI       14-APR-2000
DEFINITION  Homo sapiens gene for bilirubin UDP-glucuronosyltransferase 1,
            promoter region and partial cds.
ACCESSION   D87674
VERSION     D87674.1  GI:3059176
KEYWORDS    bilirubin UDP-glucuronosyltransferase 1.
SOURCE      Homo sapiens DNA.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Ueyama,H., Koiwai,O., Soeda,Y., Sato,H., Satoh,Y., Ohkubo,I. and
            Doida,Y.
  TITLE     Analysis of the promoter of human bilirubin
            UDP-glucuronosyltransferase gene (UGT1*1) in relevance to Gilbert's
            syndrome
  JOURNAL   Hepatol. Res. 9, 152-163 (1997)
REFERENCE   2  (bases 1 to 3341)
  AUTHORS   Ueyama,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-SEP-1996) to the DDBJ/EMBL/GenBank databases. Hisao
            Ueyama, Shiga University of Medical Science, Department of Medical
            Biochemistry; Seta, Otsu, Shiga 520-21, Japan (Tel:077-548-2162,
            Fax:077-548-2164)
COMMENT     Sequence updated (08-Jan-1997) by: Hisao Ueyama.
FEATURES             Location/Qualifiers
     source          1. .3341
```

```
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="2q37"
     misc_feature    2538. .2544
                     /note="XRE"
     misc_feature    2606. .2610
                     /note="XRE"
     misc_feature    3088. .3097
                     /note="E-box"
     misc_feature    3101. .3113
                     /note="HNF-1 site"
     CAAT_signal     3125. .3129
     TATA_signal     3149. .3153
     exon            3177. .3341
                     /number=1
     gene            3192. .3341
                     /gene="UGT1*1"
     CDS             3192. .>3341
                     /gene="UGT1*1"
                     /codon_start=1
                     /product="bilirubin UDP-glucuronosyltransferase 1"
                     /protein_id="BAA25600.1"
                     /db_xref="GI:3059177"
                     /translation="MAVESQGGRPLVLGLLLCVLGPVVSHAGKILLIPVDGSHWLSML
                     GAIQQL"
BASE COUNT      893 a    695 c    803 g    950 t
ORIGIN


  Query Match           100.0%;  Score 21;  DB 91;  Length 3341;
  Best Local Similarity  100.0%;  Pred. No. 0.73;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 aagtgaactccctgctacctt 21
          |||||||||||||||||||||
Db   3045 AAGTGAACTCCCTGCTACCTT 3065


RESULT   9
AC006985/c
LOCUS       AC006985    68770 bp    DNA             PRI       21-DEC-1999
DEFINITION  Homo sapiens BAC clone RP11-154L24 from 2, complete sequence.
ACCESSION   AC006985
VERSION     AC006985.2  GI:5732165
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 68770)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
  MEDLINE   99063792
REFERENCE   2  (bases 1 to 68770)
  AUTHORS   Gattung,S., Stoneking,T. and Davidson,T.
  TITLE     The sequence of Homo sapiens BAC clone RP11-154L24
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 68770)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-MAR-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 68770)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-AUG-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   5  (bases 1 to 68770)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-OCT-1999) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   6  (bases 1 to 68770)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-DEC-1999) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Aug 13, 1999 this sequence version replaced gi:4337256.
            ---------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu/gsc
                Contact: sapiens@watson.wustl.edu
            ---------- Summary Statistics
                Center project name: H_NH0154L24
            ----------.

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            Mapping information for this clone was provided by Dr. John D.
            McPherson, Department of Genetics, Washington University, St. Louis
            MO.  For additional information about the map position of this
            sequence, see http://genome.wustl.edu/gsc

            SOURCE INFORMATION:
            The RPCI-11 human BAC library was made from the blood of one male
            donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
            Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
            approach for construction of bacterial artificial chromosome
            libraries. Genomics 51:1-8.  The clone may be obtained either from
            Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
            and coworkers at the Roswell Park Cancer Institute
            (http://bacpac.med.buffalo.edu)
            VECTOR:  pBACe3.6
            NEIGHBORING SEQUENCE INFORMATION:
            The clone sequenced to the left is RP11-332L11, 200 bp overlap.
            Actual start of this clone is at base position 86134 of
            RP11-332L11; actual end is at base position 68770 of RP11-154L24.

            The clone RP11-154L24 contains a tandem repeat from base positions
            38234 to 39039, this region contains some low quality data. The
            assembly is consistent with the restriction digest information.
FEATURES             Location/Qualifiers
     source          1. .68770
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /map="2"
                     /clone="RP11-154L24"
                     /clone_lib="RPCI-11"
     repeat_region   3. .192
                     /rpt_family="L1"
     repeat_region   221. .528
                     /rpt_family="Alu"
     repeat_region   529. .1634
                     /rpt_family="L1"
     repeat_region   1658. .1878
                     /rpt_family="MIR"
     repeat_region   2079. .2271
                     /rpt_family="MER1_type"
     repeat_region   2276. .2336
```

```
                     /rpt_family="MER1_type"
     repeat_region   4683. .4806
                     /rpt_family="MIR"
     repeat_region   5131. .5389
                     /rpt_family="MaLR"
     repeat_region   8738. .8871
                     /rpt_family="MIR"
     repeat_region   9092. .9371
                     /rpt_family="AcHobo"
     repeat_region   9967. .10582
                     /rpt_family="L1"
     repeat_region   10576. .11200
                     /rpt_family="L1"
     repeat_region   11196. .11469
                     /rpt_family="L1"
     repeat_region   11540. .11654
                     /rpt_family="L1"
     repeat_region   12108. .12523
                     /rpt_family="MaLR"
     repeat_region   13113. .13149
                     /rpt_family="MER1_type"
     repeat_region   13150. .13171
                     /rpt_family="(CAG)n"
     repeat_region   13172. .13286
                     /rpt_family="MER1_type"
     repeat_region   13338. .13497
                     /rpt_family="MER1_type"
     repeat_region   14727. .14848
                     /rpt_family="MIR"
     repeat_region   15890. .16244
                     /rpt_family="MaLR"
     repeat_region   16380. .16723
                     /rpt_family="L1"
     repeat_region   17483. .17560
                     /rpt_family="L2"
     repeat_region   17655. .17747
                     /rpt_family="L2"
     repeat_region   18220. .18337
                     /rpt_family="MER1_type"
     repeat_region   18770. .18816
                     /rpt_family="(CA)n"
     repeat_region   19103. .19121
                     /rpt_family="AT_rich"
     repeat_region   19122. .19636
                     /rpt_family="L1"
     repeat_region   19637. .19662
                     /rpt_family="AT_rich"
     repeat_region   19825. .20014
                     /rpt_family="L1"
     repeat_region   20034. .20098
                     /rpt_family="MIR"
     repeat_region   20254. .21067
                     /rpt_family="L1"
     repeat_region   21141. .21182
                     /rpt_family="A-rich"
     repeat_region   24649. .25018
                     /rpt_family="Retroviral"
     repeat_region   25258. .25447
                     /rpt_family="MIR"
     repeat_region   25777. .25824
                     /rpt_family="(CA)n"
     repeat_region   28020. .28178
                     /rpt_family="MIR"
     repeat_region   28178. .28222
                     /rpt_family="L2"
     repeat_region   28503. .28558
                     /rpt_family="L1"
     repeat_region   28558. .28741
                     /rpt_family="L1"
     repeat_region   28970. .29064
                     /rpt_family="L1"
     repeat_region   29276. .29434
                     /rpt_family="MER1_type"
     repeat_region   29554. .29712
                     /rpt_family="MER1_type?"
     repeat_region   30867. .31015
                     /rpt_family="MER1_type"
     repeat_region   31268. .31682
                     /rpt_family="MaLR"
     repeat_region   32195. .32390
                     /rpt_family="MIR"
     repeat_region   33607. .33760
                     /rpt_family="L2"
     repeat_region   33835. .34127
                     /rpt_family="Alu"
     repeat_region   34167. .34283
                     /rpt_family="GA-rich"
     repeat_region   34806. .34976
                     /rpt_family="L2"
     repeat_region   35247. .35719
                     /rpt_family="L2"
     repeat_region   35720. .35762
                     /rpt_family="GA-rich"
     repeat_region   35763. .35909

  Query Match           100.0%;  Score 21;  DB 86;  Length 68770;
  Best Local Similarity  100.0%;  Pred. No. 0.62;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 aagtgaactccctgctacctt 21
           |||||||||||||||||||||
Db   66302 AAGTGAACTCCCTGCTACCTT 66282


RESULT  10
AC026497/c
LOCUS       AC026497   176619 bp    DNA             HTG       22-MAR-2000
DEFINITION  Homo sapiens chromosome 11 clone RP11-689A10 map 11, *** SEQUENCING
            IN PROGRESS ***, 43 unordered pieces.
ACCESSION   AC026497
VERSION     AC026497.1  GI:7283186
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 176619)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 11, clone RP11-689A10
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 176619)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-MAR-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     All repeats were identified using RepeatMasker:
```

```
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
-------------- Genome Center
    Center: Whitehead Institute/ MIT Center for Genome Research
    Center code: WIBR
    Web site: http://www-seq.wi.mit.edu
    Contact: sequence_submissions@genome.wi.mit.edu
-------------- Project Information
    Center project name: L6277
    Center clone name: 689_A_10
--------------.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 43 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*       1     1333: contig of 1333 bp in length
*    1334 1433: gap of      100 bp
*    1434     2787: contig of 1354 bp in length
*    2788 2887: gap of      100 bp
*    2888     3998: contig of 1111 bp in length
*    3999 4098: gap of      100 bp
*    4099     5777: contig of 1679 bp in length
*    5778 5877: gap of      100 bp
*    5878     7685: contig of 1808 bp in length
*    7686 7785: gap of      100 bp
*    7786     8176: contig of 391 bp in length
*    8177 8276: gap of      100 bp
*    8277     9926: contig of 1650 bp in length
*    9927 10026: gap of      100 bp
*   10027    11499: contig of 1473 bp in length
*   11500 11599: gap of      100 bp
*   11600    13616: contig of 2017 bp in length
*   13617 13716: gap of      100 bp
*   13717    16073: contig of 2357 bp in length
*   16074 16173: gap of      100 bp
*   16174    18510: contig of 2337 bp in length
*   18511 18610: gap of      100 bp
*   18611    20495: contig of 1885 bp in length
*   20496 20595: gap of      100 bp
*   20596    23548: contig of 2953 bp in length
*   23549 23648: gap of      100 bp
*   23649    25910: contig of 2262 bp in length
*   25911 26010: gap of      100 bp
*   26011    28060: contig of 2050 bp in length
*   28061 28160: gap of      100 bp
*   28161    30134: contig of 1974 bp in length
*   30135 30234: gap of      100 bp
*   30235    32646: contig of 2412 bp in length
*   32647 32746: gap of      100 bp
*   32747    35096: contig of 2350 bp in length
*   35097 35196: gap of      100 bp
*   35197    37696: contig of 2500 bp in length
*   37697 37796: gap of      100 bp
*   37797    40539: contig of 2743 bp in length
*   40540 40639: gap of      100 bp
*   40640    45518: contig of 4879 bp in length
*   45519 45618: gap of      100 bp
*   45619    48728: contig of 3110 bp in length
*   48729 48828: gap of      100 bp
*   48829    52099: contig of 3271 bp in length
*   52100 52199: gap of      100 bp
*   52200    56747: contig of 4548 bp in length
*   56748 56847: gap of      100 bp
*   56848    60467: contig of 3620 bp in length
*   60468 60567: gap of      100 bp
*   60568    64947: contig of 4380 bp in length
*   64948 65047: gap of      100 bp
*   65048    69431: contig of 4384 bp in length
*   69432 69531: gap of      100 bp
*   69532    74167: contig of 4636 bp in length
*   74168 74267: gap of      100 bp
*   74268    78188: contig of 3921 bp in length
*   78189 78288: gap of      100 bp
*   78289    81664: contig of 3376 bp in length
*   81665 81764: gap of      100 bp
*   81765    85169: contig of 3405 bp in length
*   85170 85269: gap of      100 bp
*   85270    90445: contig of 5176 bp in length
*   90446 90545: gap of      100 bp
*   90546    95008: contig of 4463 bp in length
*   95009 95108: gap of      100 bp
*   95109   100587: contig of 5479 bp in length
*  100588 100687: gap of      100 bp
*  100688   105884: contig of 5197 bp in length
*  105885 105984: gap of      100 bp
*  105985   111486: contig of 5502 bp in length
*  111487 111586: gap of      100 bp
*  111587   117781: contig of 6195 bp in length
*  117782 117881: gap of      100 bp
*  117882   125343: contig of 7462 bp in length
*  125344 125443: gap of      100 bp
*  125444   132508: contig of 7065 bp in length
*  132509 132608: gap of      100 bp
*  132609   141190: contig of 8582 bp in length
*  141191 141290: gap of      100 bp
*  141291   152556: contig of 11266 bp in length
*  152557 152656: gap of      100 bp
*  152657   163590: contig of 10934 bp in length
*  163591 163690: gap of      100 bp
*  163691   176619: contig of 12929 bp in length.
FEATURES             Location/Qualifiers
     source          1. .176619
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11"
                     /clone="RP11-689A10"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1333
                     /note="assembly_fragment"
     misc_feature    1434. .2787
                     /note="assembly_fragment"
     misc_feature    2888. .3998
                     /note="assembly_fragment"
     misc_feature    4099. .5777
                     /note="assembly_fragment"
     misc_feature    5878. .7685
                     /note="assembly_fragment"
     misc_feature    7786. .8176
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:right"
     misc_feature    8277. .9926
                     /note="assembly_fragment"
     misc_feature    10027. .11499
                     /note="assembly_fragment"
     misc_feature    11600. .13616
                     /note="assembly_fragment"
     misc_feature    13717. .16073
                     /note="assembly_fragment"
     misc_feature    16174. .18510
                     /note="assembly_fragment"
     misc_feature    18611. .20495
                     /note="assembly_fragment"
     misc_feature    20596. .23548
                     /note="assembly_fragment"
     misc_feature    23649. .25910
                     /note="assembly_fragment"
     misc_feature    26011. .28060
                     /note="assembly_fragment"
     misc_feature    28161. .30134
                     /note="assembly_fragment"
```

```
     misc_feature    30235. .32646
                     /note="assembly_fragment"
     misc_feature    32747. .35096
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
     misc_feature    35197. .37696
                     /note="assembly_fragment"
     misc_feature    37797. .40539
                     /note="assembly_fragment"
     misc_feature    40640. .45518
                     /note="assembly_fragment"
     misc_feature    45619. .48728
                     /note="assembly_fragment"

  Query Match             100.0%;  Score 21;  DB 70;  Length 176619;
  Best Local Similarity   100.0%;  Pred. No. 0.59;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 aagtgaactccctgctacctt 21
          |||||||||||||||||||||
Db  89548 AAGTGAACTCCCTGCTACCTT 89528


RESULT  11
AF297093
LOCUS       AF297093   198872 bp    DNA             PRI       07-NOV-2000
DEFINITION  Homo sapiens UGT1 gene locus, complete sequence.
ACCESSION   AF297093
VERSION     AF297093.1  GI:11118740
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 198872)
  AUTHORS   Owens,I.S., Gong,Q., Cho,J.W., Huang,T., Potter,C., Gholami,N.,
            Basu,N.K., Kubota,S., Carvalho,S. and Pennington,M.W.
  TITLE     Thirteen UDP glucuronosyltransferase genes encoded at the human
            UGT1 locus
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 198872)
  AUTHORS   Owens,I.S., Gong,Q., Cho,J.W., Huang,T., Potter,C., Gholami,N.,
            Basu,N.K., Kubota,S., Carvalho,S. and Pennington,M.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-AUG-2000) Heritable Disorders Branch, NIH-NICHD, 9000
            Rockville Pike, Bethesda, MD 20892, USA
FEATURES             Location/Qualifiers
     source          1. .198872
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     gene            1989. .2841
                     /gene="UGT1A12p"
                     /note="UDP glucuronosyltransferase 1A12"
                     /pseudo
     CDS             1989. .2841
                     /gene="UGT1A12p"
                     /codon_start=1
                     /pseudo
     gene            19970. .20956
                     /gene="UGT1A11p"
                     /note="UDP glucuronosyltransferase 1A11"
                     /pseudo
     TATA_signal     19970. .19978
     CDS             20105. .20956
                     /gene="UGT1A11p"
                     /codon_start=1
                     /pseudo
     TATA_signal     34133. .34149
                     /gene="UGT1A8"
     gene            34133. .>187313
                     /gene="UGT1A8"
     mRNA            join(<34264. .35118,181788. .181919,182603. .182690,
                     182974. .183193,187016. .>187313)
                     /gene="UGT1A8"
                     /product="UDP glucuronosyltransferase 1A8"
     CDS             join(34264. .35118,181788. .181919,182603. .182690,
                     182974. .183193,187016. .187313)
                     /gene="UGT1A8"
                     /codon_start=1
                     /product="UDP glucuronosyltransferase 1A8"
                     /protein_id="AAG30416.1"
                     /db_xref="GI:11118741"
                     /translation="MARTGWTSPIPLCVSLLLTCGFAEAGKLLVVPMDGSHWFTMQSV
                     VEKLILRGHEVVVVMPEVSWQLGKSLNCTVKTYSTSYTLEDLDREFMDFADAQWKAQV
                     RSLFSLFLSSSNGFFNLFFSHCRSLFNDRKLVEYLKESSFDAVFLDPFDACGLIVAKY
                     FSLPSVVFARGIACHYLEEGAQCPAPLSYVPRILLGFSDAMTFKERVRNHIMHLEEHL
                     FCQYFSKNALEIASEILQTPVTAYDLYSHTSIWLLRTDFVLDYPKPVMPNMIFIGGIN
                     CHQGKPLPMEFEAYINASGEHGIVVFSLGSMVSEIPEKKAMAIADALGKIPQTVLWRY
                     TGTRPSNLANNTILVKWLPQNDLLGHPMTRAFITHAGSHGVYESICNGVPMVMMPLFG
                     DQMDNAKRMETKGAGVTLNVLEMTSEDLENALKAVINDKSYKENIMRLSSLHKDRPVE
                     PLDLAVFWVEFVMRHKGAPHLRPAAHDLTWYQYHSLDVIGFLLAVVLTVAFITFKCCA
                     YGYRKCLGKKGRVKKAHKSKTH"
     TATA_signal     52972. .52988
                     /gene="UGT1A10"
     gene            52972. .>187313
                     /gene="UGT1A10"
     mRNA            join(<53103. .53957,181788. .181919,182603. .182690,
                     182974. .183193,187016. .>187313)
                     /gene="UGT1A10"
                     /product="UDP glucuronosyltransferase 1A10"
     CDS             join(53103. .53957,181788. .181919,182603. .182690,
                     182974. .183193,187016. .187313)
                     /gene="UGT1A10"
                     /codon_start=1
                     /product="UDP glucuronosyltransferase 1A10"
                     /protein_id="AAG30417.1"
                     /db_xref="GI:11118742"
                     /translation="MARAGWTSPVPLCVCLLLTCGFAEAGKLLVVPMDGSHWFTMQSV
                     VEKLILRGHEVVVVMPEVSWQLERSLNCTVKTYSTSYTLEDQNREFMVFAHAQWKAQA
                     QSIFSLLMSSSSGFLDLFFSHCRSLFNDRKLVEYLKESSFDAVFLDPFDTCGLIVAKY
                     FSLPSVVFTRGIFCHHLEEGAQCPAPLSYVPNDLLGFSDAMTFKERVWNHIVHLEDHL
                     FCQYLFRNALEIASEILQTPVTAYDLYSHTSIWLLRTDFVLDYPKPVMPNMIFIGGIN
                     CHQGKPLPMEFEAYINASGEHGIVVFSLGSMVSEIPEKKAMAIADALGKIPQTVLWRY
                     TGTRPSNLANNTILVKWLPQNDLLGHPMTRAFITHAGSHGVYESICNGVPMVMMPLFG
                     DQMDNAKRMETKGAGVTLNVLEMTSEDLENALKAVINDKSYKENIMRLSSLHKDRPVE
                     PLDLAVFWVEFVMRHKGAPHLRPAAHDLTWYQYHSLDVIGFLLAVVLTVAFITFKCCA
                     YGYRKCLGKKGRVKKAHKSKTH"
     gene            64509. .65608
                     /gene="UGT1A13p"
                     /note="UDP glucuronosyltransferase 1A13"
                     /pseudo
     CDS             join(64509. .65011,65353. .65608)
                     /gene="UGT1A13p"
                     /codon_start=1
                     /pseudo
     gene            88416. .>187313
                     /gene="UGT1A9"
     TATA_signal     88416. .88427
                     /gene="UGT1A9"
     mRNA            join(<88543. .89397,181788. .181919,182603. .182690,
                     182974. .183193,187016. .>187313)
                     /gene="UGT1A9"
                     /product="UDP glucuronosyltransferase 1A9"
     CDS             join(88543. .89397,181788. .181919,182603. .182690,
                     182974. .183193,187016. .187313)
                     /gene="UGT1A9"
                     /codon_start=1
                     /product="UDP glucuronosyltransferase 1A9"
                     /protein_id="AAG30418.1"
                     /db_xref="GI:11118743"
                     /translation="MACTGWTSPLPLCVCLLLTCGFAEAGKLLVVPMDGSHWFTMRSV
                     VEKLILRGHEVVVVMPEVSWQLGRSLNCTVKTYSTSYTLEDLDREFKAFAHAQWKAQV
                     RSIYSLLMGSYNDIFDLFFSNCRSLFKDKKLVEYLKESSFDAVFLDPFDNCGLIVAKY
                     FSLPSVVFARGILCHYLEEGAQCPAPLSYVPRILLGFSDAMTFKERVRNHIMHLEEHL
                     LCHRFFKNALEIASEILQTPVTEYDLYSHTSIWLLRTDFVLDYPKPVMPNMIFIGGIN
```

```
                CHQGKPLPMEFEAYINASGEHGIVVFSLGSMVSEIPEKKAMAIADALGKIPQTVLWRY
                TGTRPSNLANNTILVKWLPQNDLLGHPMTRAFITHAGSHGVYESICNGVPMVMMPLFG
                DQMDNAKRMETKGAGVTLNVLEMTSEDLENALKAVINDKSYKENIMRLSSLHKDRPVE
                PLDLAVFWVEFVMRHKGAPHLRPAAHDLTWYQYHSLDVIGFLLAVVLTVAFITFKCCA
                YGYRKCLGKKGRVKKAHKSKTH"
TATA_signal     98424. .98438
                /gene="UGT1A7"
gene            98424. .>187313
                /gene="UGT1A7"
mRNA            join(<98552. .99406,181788. .181919,182603. .182690,
                182974. .183193,187016. .>187313)
                /gene="UGT1A7"
                /product="UDP glucuronosyltransferase 1A7"
CDS             join(98552. .99406,181788. .181919,182603. .182690,
                182974. .183193,187016. .187313)
                /gene="UGT1A7"
                /codon_start=1
                /product="UDP glucuronosyltransferase 1A7"
                /protein_id="AAG30419.1"
                /db_xref="GI:11118744"
                /translation="MARAGWTGLLPLYVCLLLTCGFAKAGKLLVVPMDGSHWFTMQSV
                VEKLILRGHEVVVVMPEVSWQLGRSLNCTVKTYSTSYTLEDQDREFMVFADARWTAPL
                RSAFSLLTSSSNGIFDLFFSNCRSLFKDKKLVEYLKESCFDAVFLDPFDACGLIVAKY
                FSLPSVVFARGIFCHYLEEGAQCPAPLSYVPRLLLGFSDAMTFKERVRNHIMHLEEHL
                FCPYFFKNVLEIASEILQTPVTAYDLYSHTSIWLLRTDFVLEYPKPVMPNMIFIGGIN
                CHQGKPVPMEFEAYINASGEHGIVVFSLGSMVSEIPEKKAMAIADALGKIPQTVLWRY
                TGTRPSNLANNTILVKWLPQNDLLGHPMTRAFITHAGSHGVYESICNGVPMVMMPLFG
                DQMDNAKRMETKGAGVTLNVLEMTSEDLENALKAVINDKSYKENIMRLSSLHKDRPVE
                PLDLAVFWVEFVMRHKGAPHLRPAAHDLTWYQYHSLDVIGFLLAVVLTVAFITFKCCA
                YGYRKCLGKKGRVKKAHKSKTH"
gene            109467. .>187313
                /gene="UGT1A6"
TATA_signal     109467. .109473
                /gene="UGT1A6"
mRNA            join(<109610. .110470,181788. .181919,182603. .182690,
                182974. .183193,187016. .>187313)
                /gene="UGT1A6"
                /product="UDP glucuronosyltransferase 1A6"
CDS             join(109610. .110470,181788. .181919,182603. .182690,
                182974. .183193,187016. .187313)
                /gene="UGT1A6"
                /codon_start=1
                /product="UDP glucuronosyltransferase 1A6"
                /protein_id="AAG30420.1"
                /db_xref="GI:11118745"
                /translation="MACLLRSFQRISAGVFFLALWGMVVGDKLLVVPQDGSHWLSMKD
                IVEVLSDRGHEIVVVVPEVNLLLKESKYYTRKIYPVPYDQEELKNRYQSFGNNHFAER
                SFLTAPQTEYRNNMIVIGLYFINCQSLLQDRDTLNFFKESKFDALFTDPALPCGVILA
                EYLGLPSVYLFRGFPCSLEHTFSRSPDPVSYIPRCYTKFSDHMTFSQRVANFLVNLLE
                PYLFYCLFSKYEELASAVLKRDVDIITLYQKVSVWLLRYDFVLEYPRPVMPNMVFIGG
                INCKKRKDLSQEFEAYINASGEHGIVVFSLGSMVSEIPEKKAMAIADALGKIPQTVLW
                RYTGTRPSNLANNTILVKWLPQNDLLGHPMTRAFITHAGSHGVYESICNGVPMVMMPL
                FGDQMDNAKRMETKGAGVTLNVLEMTSEDLENALKAVINDKSYKENIMRLSSLHKDRP
                VEPLDLAVFWVEFVMRHKGAPHLRPAAHDLTWYQYHSLDVIGFLLAVVLTVAFITFKC
                CAYGYRKCLGKKGRVKKAHKSKTH"
gene            129525. .>187313
                /gene="UGT1A5"
TATA_signal     129525. .129531
                /gene="UGT1A5"
mRNA            join(<129600. .130466,181788. .181919,182603. .182690,
                182974. .183193,187016. .>187313)
                /gene="UGT1A5"
                /product="UDP glucuronosyltransferase 1A5"
CDS             join(129600. .130466,181788. .181919,182603. .182690,
                182974. .183193,187016. .187313)
                /gene="UGT1A5"
                /codon_start=1
                /product="UDP glucuronosyltransferase 1A5"
                /protein_id="AAG30421.1"
                /db_xref="GI:11118746"
                /translation="MATGLQVPLPQLATGLLLLLSVQPWAESGKVLVVPTDGSHWLSM
                REALRDLHARGHQVVVLTLEVNMYIKEENFFTLTTYAISWTQDEFDRLLLGHTQSFFE
                TEHLLMKFSRRMAIMNNMSLIIHRSCVELLHNEALIRHLHATSFDVVLTDPFHLCAAV
                LAKYLSIPAVFFLRNIPCDLDFKGTQCPNPSSYIPRLLTTNSDHMTFLQRVKNMLYPL
                ALSYLCHAVSAPYASLASELFQREVSVVDLVSHASVWLFRGDFVMDYPRPIMPNMVFI
                GGINCANGKPLSQEFEAYINASGEHGIVVFSLGSMVSEIPEKKAMAIADALGKIPQTV
                LWRYTGTRPSNLANNTILVKWLPQNDLLGHPMTRAFITHAGSHGVYESICNGVPMVMM
                PLFGDQMDNAKRMETKGAGVTLNVLEMTSEDLENALKAVINDKSYKENIMRLSSLHKD
                RPVEPLDLAVFWVEFVMRHKGAPHLRPAAHDLTWYQYHSLDVIGFLLAVVLTVAFITF
                KCCAYGYRKCLGKKGRVKKAHKSKTH"
```

Query Match 100.0%; Score 21; DB 89; Length 198872;
Best Local Similarity 100.0%; Pred. No. 0.59;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 aagtgaactccctgctacctt 21
          |||||||||||||||||||||
Db 174896 AAGTGAACTCCCTGCTACCTT 174916
```

```
RESULT 12
AC008278/c
LOCUS       AC008278    136868 bp    DNA             PRI       07-OCT-2000
DEFINITION  Homo sapiens BAC clone RP11-422A6 from 2, complete sequence.
ACCESSION   AC008278
VERSION     AC008278.3  GI:7408138
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 136868)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
  MEDLINE   99063792
REFERENCE   2  (bases 1 to 136868)
  AUTHORS   Du,F., Stoneking,T. and Doebber,A.
  TITLE     The sequence of Homo sapiens BAC clone RP11-422A6
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 136868)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-JUL-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 136868)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-APR-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   5  (bases 1 to 136868)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-APR-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   6  (bases 1 to 136868)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-OCT-2000) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Apr 4, 2000 this sequence version replaced gi:5870295.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu/gsc
                Contact: sapiens@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: H_NH0422A06
            --------------.

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
```

between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E., Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (http://bacpac.med.buffalo.edu)
VECTOR: pBACe3.6
NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is RP11-32P22, 200 bp overlap; the clone sequenced to the right is RP11-123E16. Actual start of this clone is at base position 149634 of RP11-32P22; actual end is at base position 136868 of RP11-422A6.

FEATURES	Location/Qualifiers
source	1. .136868
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="2"
	/map="2"
	/clone="RP11-422A6"
	/clone_lib="RPCI-11"
repeat_region	1. .70
	/rpt_family="L2"
repeat_region	533. .1406
	/rpt_family="Retroviral"
repeat_region	1407. .1441
	/rpt_family="(TAAA)n"
repeat_region	1442. .1453
	/rpt_family="Retroviral"
repeat_region	2009. .2244
	/rpt_family="L2"
repeat_region	2245. .2618
	/rpt_family="MaLR"
repeat_region	2619. .2749
	/rpt_family="L2"
repeat_region	3143. .3211
	/rpt_family="GA-rich"
repeat_region	3254. .3515
	/rpt_family="L1"
repeat_region	3516. .3808
	/rpt_family="Alu"
repeat_region	3824. .4022
	/rpt_family="L1"
repeat_region	4018. .4666
	/rpt_family="L1"
repeat_region	5232. .5530
	/rpt_family="Alu"
misc_feature	5444. .5782
	/note="match to EST AA610222 (NID:g2458650) np90c01.s1"
repeat_region	7003. .7144
	/rpt_family="L2"
repeat_region	7145. .7433
	/rpt_family="Alu"
repeat_region	7434. .7478
	/rpt_family="L2"
repeat_region	7840. .8038
	/rpt_family="MIR"
repeat_region	8232. .8512
	/rpt_family="Alu"
repeat_region	8513. .8556
	/rpt_family="(TAAA)n"
repeat_region	8656. .8969
	/rpt_family="Alu"
repeat_region	9127. .9203
	/rpt_family="MIR"
repeat_region	9361. .9669
	/rpt_family="Alu"
repeat_region	9812. .9922
	/rpt_family="L2"
repeat_region	10144. .10164
	/rpt_family="AT_rich"
repeat_region	10591. .10670
	/rpt_family="MER1_type"
repeat_region	10842. .10870
	/rpt_family="AT_rich"
repeat_region	11108. .11335
	/rpt_family="MER2_type"
repeat_region	11579. .11856
	/rpt_family="Alu"
repeat_region	12329. .12363
	/rpt_family="AT_rich"
repeat_region	12469. .12692
	/rpt_family="CT-rich"
misc_feature	12848. .12923
	/note="match to EST AA247962 (NID:g1880769)"
repeat_region	13890. .14062
	/rpt_family="Alu"
repeat_region	14856. .14900
	/rpt_family="(CATA)n"
misc_feature	15568. .15604
	/note="match to EST AA247962 (NID:g1880769)"
repeat_region	16857. .16916
	/rpt_family="MER1_type"
repeat_region	16916. .17114
	/rpt_family="MER1_type"
repeat_region	17199. .17253
	/rpt_family="(TA)n"
misc_feature	17365. .17419
	/note="match to EST AA247962 (NID:g1880769)"
repeat_region	17575. .17872
	/rpt_family="Alu"
repeat_region	18341. .18450
	/rpt_family="L2"
repeat_region	18950. .19018
	/rpt_family="L2"
repeat_region	19030. .19063
	/rpt_family="(TAAAA)n"
repeat_region	19260. .19380
	/rpt_family="L2"
misc_feature	20041. .20093
	/note="match to EST AA247962 (NID:g1880769)"
repeat_region	20254. .20359
	/rpt_family="MIR"
repeat_region	20461. .20637
	/rpt_family="L2"
repeat_region	20665. .21076
	/rpt_family="MER4-group"
repeat_region	21076. .21188
	/rpt_family="MER4-group"
repeat_region	21288. .21313
	/rpt_family="(TTCA)n"
repeat_region	21490. .22051
	/rpt_family="L2"
repeat_region	22552. .22833
	/rpt_family="Alu"
misc_feature	22955. .23059
	/note="match to EST AI150474 (NID:g3678943) qf41d07.x1"
repeat_region	23322. .23450

```
  Query Match          87.6%;  Score 18.4;  DB 86;  Length 136868;
  Best Local Similarity  95.0%;  Pred. No. 17;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 aagtgaactccctgctacct 20
          |||| ||||||||||||||||
Db  20890 AAGTTAACTCCCTGCTACCT 20871


RESULT  13
AP000293
LOCUS       AP000293     7218 bp    DNA             PRI       20-NOV-1999
DEFINITION  Homo sapiens genomic DNA, chromosome 21q22.1, D21S226-AML region,
            clone:Q20D3, complete sequence.
ACCESSION   AP000293
VERSION     AP000293.1  GI:4835662
KEYWORDS    HTG.
SOURCE      Homo sapiens DNA, clone:Q20D3.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 7218)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y. and Sakaki,Y.
  TITLE     Homo sapiens 7,218bp genomic DNA of 21q22.1
  JOURNAL   Published Only in DataBase (1999) In press
REFERENCE   2  (bases 1 to 7218)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-1999) to the DDBJ/EMBL/GenBank databases.
            Masahira Hattori, The Institute of Physical and Chemical Research
            (RIKEN), Genomic Sciences Center (GSC); Kitasato Univ., 1-15-1
            Kitasato, Sagamihara, Kanagawa 228-8555, Japan
            (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-42-778-9923, Fax:81-42-778-9924)
COMMENT     The sequence is a part of the data (ACCESSION No. AP000174 -
            AP000194).
            The sequencing project is supported by Japan Science Technology
            Corporation (JST) and The Institute of Physical and Chemical
            Research (RIKEN).
FEATURES             Location/Qualifiers
     source          1. .7218
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="21"
                     /clone="Q20D3"
                     /map="21q22.1"
BASE COUNT     2067 a   1568 c   1659 g   1924 t
ORIGIN


  Query Match          84.8%;  Score 17.8;  DB 91;  Length 7218;
  Best Local Similarity  90.5%;  Pred. No. 45;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 aagtgaactccctgctacctt 21
          |||| | ||||||||||||||||
Db   3890 AAGTTAGCTCCCTGCTACCTT 3910


RESULT  14
AC087204
LOCUS       AC087204    46831 bp    DNA             HTG       13-DEC-2000
DEFINITION  Homo sapiens chromosome 8 clone RP11-520F7 map 8, LOW-PASS SEQUENCE
            SAMPLING.
ACCESSION   AC087204
VERSION     AC087204.1  GI:11693386
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 46831)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 8, clone RP11-520F7
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 46831)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B., Brown,A.,
            Camarata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J.,
            Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Karatas,A., LaRocque,K., Lamazares,R., Landers,T.,
            Lehoczky,J., Levine,R., Liu,G., MacLean,C., Macdonald,P.,
            Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,
            Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M.,
            Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Sougnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A.,
            Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-DEC-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            ---------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            ---------- Project Information
                Center project name: Ll1634
                Center clone name: 520_F_7
            ----------.
            * NOTE: This record contains 61 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *        1     676: contig of 676 bp in length
            *      677 776: gap of      100 bp
            *      777    1441: contig of 665 bp in length
            *     1442 1541: gap of      100 bp
            *     1542    2219: contig of 678 bp in length
            *     2220 2319: gap of      100 bp
            *     2320    2980: contig of 661 bp in length
            *     2981 3080: gap of      100 bp
            *     3081    3757: contig of 677 bp in length
            *     3758 3857: gap of      100 bp
            *     3858    4517: contig of 660 bp in length
            *     4518 4617: gap of      100 bp
            *     4618    5331: contig of 714 bp in length
            *     5332 5431: gap of      100 bp
            *     5432    6074: contig of 643 bp in length
            *     6075 6174: gap of      100 bp
            *     6175    6852: contig of 678 bp in length
            *     6853 6952: gap of      100 bp
            *     6953    7610: contig of 658 bp in length
            *     7611 7710: gap of      100 bp
```

```
*     7711    8355: contig of 645 bp in length
*     8356 8455: gap of     100 bp
*     8456    9129: contig of 674 bp in length
*     9130 9229: gap of     100 bp
*     9230    9891: contig of 662 bp in length
*     9892 9991: gap of     100 bp
*     9992   10665: contig of 674 bp in length
*    10666 10765: gap of     100 bp
*    10766   11423: contig of 658 bp in length
*    11424 11523: gap of     100 bp
*    11524   12180: contig of 657 bp in length
*    12181 12280: gap of     100 bp
*    12281   12945: contig of 665 bp in length
*    12946 13045: gap of     100 bp
*    13046   13692: contig of 647 bp in length
*    13693 13792: gap of     100 bp
*    13793   14442: contig of 650 bp in length
*    14443 14542: gap of     100 bp
*    14543   15203: contig of 661 bp in length
*    15204 15303: gap of     100 bp
*    15304   15968: contig of 665 bp in length
*    15969 16068: gap of     100 bp
*    16069   16747: contig of 679 bp in length
*    16748 16847: gap of     100 bp
*    16848   17509: contig of 662 bp in length
*    17510 17609: gap of     100 bp
*    17610   18256: contig of 647 bp in length
*    18257 18356: gap of     100 bp
*    18357   19030: contig of 674 bp in length
*    19031 19130: gap of     100 bp
*    19131   19779: contig of 649 bp in length
*    19780 19879: gap of     100 bp
*    19880   20538: contig of 659 bp in length
*    20539 20638: gap of     100 bp
*    20639   21294: contig of 656 bp in length
*    21295 21394: gap of     100 bp
*    21395   22069: contig of 675 bp in length
*    22070 22169: gap of     100 bp
*    22170   22845: contig of 676 bp in length
*    22846 22945: gap of     100 bp
*    22946   23627: contig of 682 bp in length
*    23628 23727: gap of     100 bp
*    23728   24395: contig of 668 bp in length
*    24396 24495: gap of     100 bp
*    24496   25150: contig of 655 bp in length
*    25151 25250: gap of     100 bp
*    25251   25913: contig of 663 bp in length
*    25914 26013: gap of     100 bp
*    26014   26661: contig of 648 bp in length
*    26662 26761: gap of     100 bp
*    26762   27441: contig of 680 bp in length
*    27442 27541: gap of     100 bp
*    27542   28201: contig of 660 bp in length
*    28202 28301: gap of     100 bp
*    28302   28991: contig of 690 bp in length
*    28992 29091: gap of     100 bp
*    29092   29766: contig of 675 bp in length
*    29767 29866: gap of     100 bp
*    29867   30521: contig of 655 bp in length
*    30522 30621: gap of     100 bp
*    30622   31300: contig of 679 bp in length
*    31301 31400: gap of     100 bp
*    31401   32074: contig of 674 bp in length
*    32075 32174: gap of     100 bp
*    32175   32854: contig of 680 bp in length
*    32855 32954: gap of     100 bp
*    32955   33639: contig of 685 bp in length
*    33640 33739: gap of     100 bp
*    33740   34430: contig of 691 bp in length
*    34431 34530: gap of     100 bp
*    34531   35185: contig of 655 bp in length
*    35186 35285: gap of     100 bp
*    35286   35958: contig of 673 bp in length
*    35959 36058: gap of     100 bp
*    36059   36729: contig of 671 bp in length
*    36730 36829: gap of     100 bp
*    36830   37509: contig of 680 bp in length
*    37510 37609: gap of     100 bp
*    37610   38263: contig of 654 bp in length
*    38264 38363: gap of     100 bp
*    38364   39039: contig of 676 bp in length
*    39040 39139: gap of     100 bp
*    39140   39807: contig of 668 bp in length
*    39808 39907: gap of     100 bp
*    39908   40578: contig of 671 bp in length
*    40579 40678: gap of     100 bp
*    40679   41357: contig of 679 bp in length
*    41358 41457: gap of     100 bp
*    41458   42167: contig of 710 bp in length
*    42168 42267: gap of     100 bp
*    42268   42963: contig of 696 bp in length
*    42964 43063: gap of     100 bp
*    43064   43713: contig of 650 bp in length
*    43714 43813: gap of     100 bp
*    43814   44501: contig of 688 bp in length
*    44502 44601: gap of     100 bp
*    44602   45262: contig of 661 bp in length
*    45263 45362: gap of     100 bp
*    45363   46040: contig of 678 bp in length
*    46041 46140: gap of     100 bp
*    46141   46831: contig of 691 bp in length.
FEATURES             Location/Qualifiers
     source          1. .46831
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /map="8"
                     /clone="RP11-520F7"
                     /clone_lib="RPCI-11 Human Male BAC"
BASE COUNT    10780 a    8764 c    9356 g   11744 t    6187 others
ORIGIN


  Query Match           84.8%;  Score 17.8;  DB 77;  Length 46831;
  Best Local Similarity  90.5%;  Pred. No. 40;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 aagtgaactccctgctacctt 21
          | |||| ||||||||||||||
Db   7326 AGGTGAGCTCCCTGCTACCTT 7346


RESULT  15
AL391843/c
LOCUS       AL391843    47972 bp    DNA             PRI       22-FEB-2001
DEFINITION  Human DNA sequence from clone RP11-592B10 on chromosome 6, complete
            sequence.
ACCESSION   AL391843
VERSION     AL391843.13  GI:13157590
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 47972)
  AUTHORS   Sycamore,N.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-FEB-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Feb 27, 2001 this sequence version replaced gi:12831877.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
```

only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping Group. Further information can be found at http://www.sanger.ac.uk/HGP/Chr6
RP11-592B10 is from the library RPCI-11.3 constructed by the group of Pieter de Jong. For further details see http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6
IMPORTANT: This sequence is not the entire insert of clone RP11-592B10 It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.
The true left end of clone RP11-411F9 is at 47873 in this sequence.
The true right end of clone RP5-1007B16 is at 100 in this sequence.

FEATURES Location/Qualifiers

source	1. .47972
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="6"
	/clone="RP11-592B10"
	/clone_lib="RPCI-11.3"
misc_feature	complement(147. .519)
	/note="match: GSS: Em:AQ253031"
misc_feature	156. .328
	/note="match: GSS: Em:AQ542267"
misc_feature	complement(156. .518)
	/note="match: GSS: Em:AQ210521"
misc_feature	complement(156. .512)
	/note="match: GSS: Em:AQ560490"
repeat_region	164. .366
	/note="LTR12 repeat: matches 18. .192 of consensus"
misc_feature	332. .503
	/note="match: GSS: Em:AQ542267"
repeat_region	427. .472
	/note="LTR12 repeat: matches 189. .234 of consensus"
repeat_region	473. .549
	/note="LTR30 repeat: matches 544. .620 of consensus"
repeat_region	480. .583
	/note="MER61E repeat: matches 377. .477 of consensus"
repeat_region	584. .673
	/note="LTR12 repeat: matches 582. .671 of consensus"
repeat_region	674. .7342
	/note="HERVK22 repeat: matches 100. .6836 of consensus"
repeat_region	7344. .7825
	/note="LTR22 repeat: matches 1. .509 of consensus"
repeat_region	7826. .11300
	/note="L1MA1 repeat: matches 685. .4254 of consensus"
misc_feature	10015. .10095
	/note="Single clone region. Assembly confirmed by restriction digest data."
repeat_region	11301. .11611
	/note="AluY repeat: matches 1. .309 of consensus"
repeat_region	11612. .13713
	/note="L1MA1 repeat: matches 4254. .6304 of consensus"
repeat_region	13766. .15206
	/note="L1M3 repeat: matches 2143. .3546 of consensus"
repeat_region	15207. .15512
	/note="AluSx repeat: matches 2. .308 of consensus"
repeat_region	15513. .17628
	/note="L1M3 repeat: matches 3546. .5738 of consensus"
repeat_region	18106. .18525
	/note="Tigger2a repeat: matches 23. .433 of consensus"
repeat_region	19408. .20039
	/note="LTR9 repeat: matches 6. .625 of consensus"
repeat_region	20213. .20397
	/note="L1M4 repeat: matches 4454. .4646 of consensus"
repeat_region	20956. .21581
	/note="MER41B repeat: matches 1. .635 of consensus"
repeat_region	22205. .22776
	/note="MER77 repeat: matches 13. .629 of consensus"
misc_feature	23549. .24036
	/note="match: GSS: Em:AQ668040"
repeat_region	23847. .23903
	/note="L2 repeat: matches 2639. .2695 of consensus"
misc_feature	complement(24422. .24864)
	/note="match: GSS: Em:AQ610704"
misc_feature	complement(24640. .24857)
	/note="match: GSS: Em:AQ030189"
repeat_region	24874. .24928
	/note="L2 repeat: matches 2692. .2746 of consensus"
misc_feature	complement(25273. .25508)
	/note="match: GSS: Em:AQ376698"
misc_feature	complement(25297. .25473)
	/note="match: GSS: Em:AQ545471"
misc_feature	complement(25299. .25496)
	/note="match: GSS: Em:AQ749890"
misc_feature	complement(25327. .25490)
	/note="match: GSS: Em:B56335"
repeat_region	26210. .26253
	/note="22 copies 2 mer tg 95% conserved"
misc_feature	26543. .26868
	/note="match: GSS: Em:AQ902789"
repeat_region	26950. .27010
	/note="MER90 repeat: matches 3. .63 of consensus"
repeat_region	27054. .27374
	/note="MER90 repeat: matches 222. .543 of consensus"
repeat_region	27375. .27869
	/note="LTR15 repeat: matches 1. .489 of consensus"
repeat_region	27870. .27942
	/note="MER90 repeat: matches 543. .615 of consensus"
repeat_region	28722. .28935
	/note="AluY repeat: matches 74. .286 of consensus"
misc_feature	complement(28723. .29206)
	/note="match: GSS: Em:AQ882343"
repeat_region	30251. .30554
	/note="AluY repeat: matches 1. .305 of consensus"
repeat_region	33533. .33906
	/note="MLT1A2 repeat: matches 1. .374 of consensus"
repeat_region	34190. .34587
	/note="MLT1J repeat: matches 117. .513 of consensus"
repeat_region	34664. .35066
	/note="MSTA repeat: matches 1. .424 of consensus"
repeat_region	35260. .35321
	/note="31 copies 2 mer aa 80% conserved"
repeat_region	35346. .35777
	/note="LTR1 repeat: matches 1. .437 of consensus"
repeat_region	35952. .36169
	/note="MLT1A1 repeat: matches 1. .210 of consensus"
repeat_region	36201. .36250
	/note="25 copies 2 mer ca 82% conserved"
repeat_region	36271. .36393
	/note="FLAM_A repeat: matches 2. .124 of consensus"
repeat_region	37540. .37613
	/note="37 copies 2 mer aa 67% conserved"
repeat_region	38348. .38399
	/note="26 copies 2 mer at 84% conserved"
misc_feature	complement(39249. .40015)
	/note="match: GSS: Em:AQ538237"
misc_feature	40021. .40482
	/note="match: GSS: Em:AQ171081"
repeat_region	40101. .40352
	/note="MIR repeat: matches 1. .259 of consensus"
repeat_region	40581. .40676
	/note="48 copies 2 mer aa 64% conserved"

```
     repeat_region   41234. .41662
                     /note="L1M2 repeat: matches -676. .-231 of consensus"
     repeat_region   42139. .42439
                     /note="L1PA15-16 repeat: matches -265. .24 of consensus"
     repeat_region   42901. .43464
                     /note="L1PA13 repeat: matches 802. .1288 of consensus"
     repeat_region   43465. .43687
                     /note="MER30 repeat: matches 1. .230 of consensus"
     repeat_region   43688. .43754
                     /note="L1PA13 repeat: matches 1288. .1351 of consensus"
     repeat_region   43755. .46738
                     /note="L1PA16 repeat: matches 3178. .6153 of consensus"
     repeat_region   46921. .47311
                     /note="MLT2B repeat: matches 1. .304 of consensus"
     misc_feature    complement(47070. .47394)
                     /note="match: GSS: Em:AQ053386"
     misc_feature    complement(47357. .47846)
                     /note="match: GSS: Em:AQ534016"
     misc_feature    47878. .47972
                     /note="match: GSS: Em:AQ186789"
     misc_feature    47880. .47972
                     /note="match: GSS: Em:AQ551896"
BASE COUNT    16896 a   8982 c   8503 g  13591 t
ORIGIN


  Query Match             84.8%;  Score 17.8;  DB 90;  Length 47972;
  Best Local Similarity   90.5%;  Pred. No. 40;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 aagtgaactccctgctaccttt 21
           ||||||||||||| || ||||
Db   29145 AAGTGAACTCCCTCCTGCCTT 29125


Search completed: July 25, 2001, 05:16:42
Job time: 9205 sec
```

OM nucleic - nucleic search, using sw model

Run on: July 25, 2001, 04:05:07 ; Search time 247.87 Seconds
(without alignments)
53.197 Million cell updates/sec

Title: US-09-142-095-1
Perfect score: 21
Sequence: 1 aagtgaactccctgctacctt 21

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_0601:*
1: /SIDS1/gcgdata/geneseq/geneseqn/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	21	100.0	21	18	AAT79541	UGT1*1 gene exon 1
	2	21	100.0	620	18	AAT79540	Upstream DNA seque
	3	16.8	80.0	237326	19	AAV57903	Hereditary haemoch
c	4	16.2	77.1	260	21	AAC25716	Human secreted pro
c	5	16.2	77.1	867	21	AAC38612	Arabidopsis thalia
c	6	16.2	77.1	1613	21	AAA46500	cDNA encoding enzy
	7	16.2	77.1	32042	20	AAZ09252	Human CARD-4 DNA.
	8	16.2	77.1	32042	22	AAF30011	Human CARD-4 gene.
	9	15.8	75.2	957	16	AAQ97865	TRONP1 DNA. Trepo
	10	15.8	75.2	957	18	AAT76580	Treponema pallidum
c	11	15.2	72.4	537	21	AAA67251	Eucalyptus grandis
	12	15.2	72.4	640	18	AAV04426	Human calpain smal
	13	15.2	72.4	804	18	AAV04420	Human calpain smal
	14	15.2	72.4	1242	21	AAC77765	Human cancer assoc
c	15	15.2	72.4	1741	21	AAZ98118	Human signal pepti
c	16	15.2	72.4	1938	21	AAC46144	Arabidopsis thalia
c	17	15.2	72.4	1940	21	AAC39039	Arabidopsis thalia
	18	15.2	72.4	2024	21	AAC81718	Human secreted pro
	19	15.2	72.4	2582	20	AAV99562	Soybean lysine ket
	20	15.2	72.4	4044	21	AAC81899	A. thaliana SRp30
	21	15.2	72.4	4665	18	AAT97955	Human epididymis-s
	22	15.2	72.4	235033	19	AAV57926	Hereditary haemoch
	23	15	71.4	1330	21	AAC33777	Arabidopsis thalia
c	24	15	71.4	14063	20	AAX20500	Polynucleotide seq
c	25	14.8	70.5	307	21	AAC98243	Human colon cancer
	26	14.8	70.5	413	21	AAC26942	Human secreted pro
c	27	14.8	70.5	731	16	AAT20482	Human gene signatu
c	28	14.8	70.5	1398	20	AAX25083	Potato tuber-speci
	29	14.8	70.5	1723	17	AAT42133	Ascorbate-free-rad
	30	14.8	70.5	2161	20	AAZ30444	Tobacco anthranila
	31	14.8	70.5	2526	20	AAZ20026	Human rad17 cell c
	32	14.8	70.5	2535	20	AAZ20033	Human rad17 cell c
	33	14.8	70.5	2540	20	AAZ20052	Human rad17 cell c
	34	14.8	70.5	2543	20	AAZ20051	Human rad17 cell c
	35	14.8	70.5	2647	20	AAZ20035	Human rad17 cell c
	36	14.8	70.5	2652	17	AAT34655	Testis-associated
c	37	14.8	70.5	2977	17	AAT30085	Human Raf1 kinase.
c	38	14.8	70.5	2977	18	AAT61894	Human raf-1 cDNA.
c	39	14.8	70.5	2977	19	AAV20439	Human c-raf oncoge
c	40	14.8	70.5	2977	20	AAX99340	Human c-raf sequen
c	41	14.8	70.5	2977	20	AAX78137	Human c-Raf-1 gene
c	42	14.8	70.5	2977	21	AAA73552	Polynucleotide #1
c	43	14.8	70.5	2977	21	AAA48654	Human c-raf DNA.
c	44	14.8	70.5	2977	22	AAF75126	Human c-Raf cDNA.
c	45	14.8	70.5	5907	21	AAC76910	Human ORFX ORF2465

ALIGNMENTS

RESULT 1
AAT79541
ID AAT79541 standard; DNA; 21 BP.
XX
AC AAT79541;
XX
DT 23-JAN-1998 (first entry)
XX
DE UGT1*1 gene exon 1 upstream PCR primer A.
XX
KW Uridine diphosphate glucuronosyltransferase gene; UGT;
KW Gilbert's syndrome; GS; unconjugated hyperbilirubinaemia;
KW bilirubin glucuronidation; Crigler-Najjar; type 2; drug metabolism;
KW Drug trial efficiency; screening; PCR primer; ss.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN WO9732042-A2.
XX
PD 04-SEP-1997.
XX
PF 03-MAR-1997; 97WO-GB00577.
XX
PR 16-MAR-1996; 96GB-0005598.
PR 01-MAR-1996; 96GB-0004480.
XX
PA (UYDU-) UNIV DUNDEE.
XX
PI Burchell B;
XX
DR WPI; 1997-448702/41.
XX
PT Improving drug trial efficiency comprises identifying participants

```
PT   with Gilbert's syndrome - useful as their altered drug metabolism
PT   may hinder result interpretation
XX
PS   Claim 14; Page 12; 31pp; English.
XX
CC   This PCR primer (with primers AAT79542-44) flanks the TATA box sequence
CC   upstream of the uridine diphosphate glucuronosyltransferase (UGT) gene
CC   1*1 exon 1 (see AAT79540), and was used to amplify fragments of 253-255
CC   bp. This gene is known to be associated with Gilbert's syndrome (GS). GS
CC   is a mild, common form of unconjugated hyperbilirubinaemia associated
CC   with reduced bilirubin glucuronidation capacity. Analysis of the genetic
CC   basis of GS has allowed 2 forms to be identified. One is a mild form
CC   associated with a homozygous 2 bp insertion in the TATA sequence
CC   upstream of the UGT1*1 exon 1, and the other is a more severe form
CC   associated with heterozygosity for a mutation which, when homozygous,
CC   causes Crigler-Najjar type 2 disease. The first form is autosomal
CC   recessive and the second is inherited dominantly. Patients suffering from
CC   GS, which is benign, may have altered metabolism of some drugs, making it
CC   difficult to determine if an effect is due to the drug or the syndrome.
CC   Drug trial efficiency would be improved if potential participants can be
CC   screened for the genetic basis of GS, and eliminated or included  on
CC   basis of them possessing or not possessing GS.
XX
SQ   Sequence 21 BP; 5 A; 7 C; 3 G; 6 T; 0 other;


  Query Match            100.0%;  Score 21;  DB 18;  Length 21;
  Best Local Similarity  100.0%;  Pred. No. 0.14;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 aagtgaactccctgctacctt 21
          |||||||||||||||||||||
Db      1 aagtgaactccctgctacctt 21


RESULT   2
AAT79540
ID   AAT79540 standard; DNA; 620 BP.
XX
AC   AAT79540;
XX
DT   23-JAN-1998  (first entry)
XX
DE   Upstream DNA sequence of UGT1*1 gene exon 1.
XX
KW   Uridine diphosphate glucuronosyltransferase gene; UGT;
KW   Gilbert's syndrome; GS; unconjugated hyperbilirubinaemia;
KW   bilirubin glucuronidation; Crigler-Najjar; type 2; drug metabolism;
KW   Drug trial efficiency; screening; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   protein_bind     2..8
FT                    /*tag=  a
FT                    /bound_moeity= AP1
FT   GC_signal        93..102
FT                    /*tag=  b
FT                    /note= "SP1 binding site"
FT   misc_feature     180..191
FT                    /*tag=  c
FT                    /note= "feature indicated in patent, but no further
FT                          explanation is given"
FT   protein_bind     316..323
FT                    /*tag=  d
FT                    /bound_moeity= AP3
FT   protein_bind     317..324
FT                    /*tag=  e
FT                    /bound_moeity= CLBP
FT   misc_feature     366..376
FT                    /*tag=  f
FT                    /note= "feature indicated in patent, but no further
FT                          explanation is given"
FT   protein_bind     386..392
FT                    /*tag=  g
FT                    /bound_moeity= AP1
FT   protein_bind     513..519
FT                    /*tag=  h
FT                    /bound_moeity= AP1
FT   misc_feature     513..520
FT                    /*tag=  i
FT                    /note= "feature indicated in patent, but no further
FT                          explanation is given"
FT   protein_bind     520..532
FT                    /*tag=  j
FT                    /bound_moeity= HNF1
FT   TATA_signal      558..572
FT                    /*tag=  k
FT                    /note= "corresponds to positions -53 to -39 in patent"
FT   misc_feature     595
FT                    /*tag=  l
FT                    /note= "feature indicated in patent, but no further
FT                          explanation is given"
FT   CDS              612
FT                    /*tag=  m
FT                    /codon_start= 612
XX
PN   WO9732042-A2.
XX
PD   04-SEP-1997.
XX
PF   03-MAR-1997;   97WO-GB00577.
XX
PR   16-MAR-1996;   96GB-0005598.
PR   01-MAR-1996;   96GB-0004480.
XX
PA   (UYDU-) UNIV DUNDEE.
XX
PI   Burchell B;
XX
DR   WPI; 1997-448702/41.
XX
PT   Improving drug trial efficiency comprises identifying participants
PT   with Gilbert's syndrome - useful as their altered drug metabolism
PT   may hinder result interpretation
XX
PS   Claim 11; Fig 4; 31pp; English.
XX
CC   This sequence represents the upstream sequence, positions -611 to 9
CC   in the patent, of uridine diphosphate glucuronosyltransferase (UGT)
CC   gene 1*1 exon 1. This gene is known to be associated with Gilbert's
CC   syndrome (GS). GS is a mild, common form of unconjugated
CC   hyperbilirubinaemia associated with reduced bilirubin glucuronidation
CC   capacity. Analysis of the genetic basis of GS has allowed 2 forms to be
CC   identified. One is a mild form associated with a homozygous 2 bp
CC   insertion in the TATA sequence upstream of the UGT1*1 exon 1, and
CC   the other is a more severe form associated with heterozygosity for a
CC   mutation which, when homozygous, causes Crigler-Najjar type 2 disease.
CC   The first form is autosomal recessive and the second is inherited
CC   dominantly. Patients suffering from GS, which is benign, may have
CC   altered metabolism of some drugs, making it difficult to determine if an
CC   effect is due to the drug or the syndrome. Drug trial efficiency would
CC   be improved if potential participants can be screened for the genetic
CC   basis of GS, and eliminated or included  on basis of them possessing or
CC   not possessing GS. In this case, screening involves PCR amplification
CC   of the UGT gene, using the primers described in AAT79541-44.
XX
SQ   Sequence 620 BP; 157 A; 127 C; 151 G; 185 T; 0 other;


  Query Match            100.0%;  Score 21;  DB 18;  Length 620;
  Best Local Similarity  100.0%;  Pred. No. 0.21;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 aagtgaactccctgctacctt 21
          |||||||||||||||||||||
```

```
Db    465 aagtgaactccctgctacctt 485


RESULT   3
AAV57903
ID   AAV57903 standard; DNA; 237326 BP.
XX
AC   AAV57903;
XX
DT   21-DEC-1998  (first entry)
XX
DE   Hereditary haemochromatosis subregion from an HH affected individual.
XX
KW   Bovine butyrophilin; BT; human hereditary haemochromatosis; HFE;
KW   diagnosis; iron metabolism; NPT3; NPT4; RoRet; BTF1; BTF2; BTF3;
KW   BTF4; BTF5; milk protein; lupus; Sjogren's syndrome; hypophosphatemia;
KW   type 1 sodium transport gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO9814466-A1.
XX
PD   09-APR-1998.
XX
PF   30-SEP-1997;   97WO-US17658.
XX
PR   07-MAY-1997;   97US-0852495.
PR   01-OCT-1996;   96US-0724394.
XX
PA   (PROG-) PROGENTIOR INC.
XX
PI   Feder JN,  Kronmal GS,  Lauer PM,  Ruddy DA,  Thomas WJ;
PI   Tsuchihashi Z,  Wolff RK;
XX
DR   WPI; 1998-240014/21.
XX
PT   Hereditary haemochromatosis gene products - used to develop products
PT   for the diagnosis and treatment of hereditary disorders in iron
PT   metabolism
XX
PS   Claim 1; Fig 9; 209pp; English.
XX
CC   The present invention describes hereditary haemochromatosis gene
CC   products from the human haemochromatosis gene. The present sequence
CC   represents a hereditary haemochromatosis subregion from an hereditary
CC   haemochromatosis (HH) affected individual. Also described is a
CC   method to determine the presence or absence of the common hereditary
CC   haemochromatosis (HFE) gene mutation in an individual comprising:
CC   (a) providing DNA or RNA from the individual; and (b) assessing the
CC   DNA or RNA for the presence or absence of a haplotype or genotype where
CC   the presence or absence of the haplotype genotype indicates the likely
CC   presence of the HFE gene mutation in the genome of the individual. The
CC   HFE gene sequences from the present invention can be used to develop
CC   products for use in the diagnosis and treatment of HFE. The present
CC   invention also describes BTF genes, which are homologues of the milk
CC   protein butyrophilin (BT), and can be used in the production of agonists
CC   and antagonists of BT function. Also described are: (1) a RoRet gene
CC   which can be used to develop products for the study, diagnosis and
CC   treatment of lupus and Sjogren's syndrome; and (2) NPT3 and NPT4 genes
CC   which are homologues of a type 1 sodium transport gene, and can
CC   similarly be used for hypophosphatemia.
XX
SQ   Sequence 237326 BP; 69596 A; 48904 C; 48217 G; 70609 T; 0 other;


  Query Match           80.0%;  Score 16.8;  DB 19;  Length 237326;
  Best Local Similarity  90.0%;  Pred. No. 53;
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 aagtgaactccctgctacct 20
          |||||||||| |||| ||||
Db 105327 aagtgaactcactgcaacct 105346


RESULT   4
AAC25716/c
ID   AAC25716 standard; cDNA; 260 BP.
XX
AC   AAC25716;
XX
DT   06-OCT-2000  (first entry)
XX
DE   Human secreted protein 5' EST, SEQ ID NO: 29791.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW   gene therapy; chromosome mapping; ss.
XX
OS   Homo sapiens.
XX
PN   EP1033401-A2.
XX
PD   06-SEP-2000.
XX
PF   21-FEB-2000; 2000EP-0200610.
XX
PR   26-FEB-1999;   99US-0122487.
XX
PA   (GEST ) GENSET.
XX
PI   Dumas Milne Edwards J,  Duclert A,  Giordano J;
XX
DR   WPI; 2000-500381/45.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS   Claim 1; SEQ ID 29791; 71pp + CD-ROM; English.
XX
CC   The present sequence is one of a large number of 5' ESTs derived from
CC   mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC   identified within the present sequence. The 5' ESTs were prepared from
CC   total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC   sequences usually correspond mainly to the 3' untranslated region (UTR)
CC   of the mRNA because they are often obtained from oligo-dT primed cDNA
CC   libraries. Such ESTs are not well suited for isolating cDNA sequences
CC   derived from the 5' ends of mRNAs and even in those cases where longer
CC   cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC   5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
CC   used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
CC   in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC   They are used to obtain upstream regulatory sequences and to design
CC   expression and secretion vectors.
XX
SQ   Sequence 260 BP; 74 A; 58 C; 45 G; 82 T; 1 other;


  Query Match           77.1%;  Score 16.2;  DB 21;  Length 260;
  Best Local Similarity  85.7%;  Pred. No. 46;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 aagtgaactccctgctacctt 21
          ||| |||||| || |||||||
Db     68 AAGGGAACTCGCTTCTACCTT 48


RESULT   5
AAC38612/c
ID   AAC38612 standard; DNA; 867 BP.
XX
AC   AAC38612;
XX
DT   17-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana DNA fragment SEQ ID NO: 21606.
XX
```

KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.

```
PR   01-SEP-1999;   99US-0151930.
PR   07-SEP-1999;   99US-0152363.
PR   10-SEP-1999;   99US-0153070.
PR   13-SEP-1999;   99US-0153758.
PR   15-SEP-1999;   99US-0154018.
PR   16-SEP-1999;   99US-0154039.
PR   20-SEP-1999;   99US-0154779.
PR   22-SEP-1999;   99US-0155139.
PR   23-SEP-1999;   99US-0155486.
PR   24-SEP-1999;   99US-0155659.
PR   28-SEP-1999;   99US-0156458.
PR   29-SEP-1999;   99US-0156596.
PR   04-OCT-1999;   99US-0157117.
PR   05-OCT-1999;   99US-0157753.
PR   06-OCT-1999;   99US-0157865.
PR   07-OCT-1999;   99US-0158029.
PR   08-OCT-1999;   99US-0158232.
PR   12-OCT-1999;   99US-0158369.
PR   13-OCT-1999;   99US-0159293.
PR   13-OCT-1999;   99US-0159294.
PR   13-OCT-1999;   99US-0159295.
PR   14-OCT-1999;   99US-0159329.
PR   14-OCT-1999;   99US-0159330.
PR   14-OCT-1999;   99US-0159331.
PR   14-OCT-1999;   99US-0159637.
PR   14-OCT-1999;   99US-0159638.
PR   18-OCT-1999;   99US-0159584.
PR   21-OCT-1999;   99US-0160741.
PR   21-OCT-1999;   99US-0160767.
PR   21-OCT-1999;   99US-0160768.
PR   21-OCT-1999;   99US-0160770.
PR   21-OCT-1999;   99US-0160814.
PR   21-OCT-1999;   99US-0160815.
PR   22-OCT-1999;   99US-0160980.
PR   22-OCT-1999;   99US-0160981.
PR   22-OCT-1999;   99US-0160989.
PR   25-OCT-1999;   99US-0161404.
PR   25-OCT-1999;   99US-0161405.
PR   25-OCT-1999;   99US-0161406.
PR   26-OCT-1999;   99US-0161359.
PR   26-OCT-1999;   99US-0161360.
PR   26-OCT-1999;   99US-0161361.
PR   28-OCT-1999;   99US-0161920.
PR   28-OCT-1999;   99US-0161992.
PR   28-OCT-1999;   99US-0161993.
PR   29-OCT-1999;   99US-0162142.

  Query Match             77.1%;  Score 16.2;  DB 21;  Length 867;
  Best Local Similarity   85.7%;  Pred. No. 53;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 aagtgaactccctgctacctt 21
          ||||||||||| ||| || ||
Db    678 AAGTGAACTCCTTGCGACTTT 658


RESULT   6
AAA46500/c
ID   AAA46500 standard; cDNA; 1613 BP.
XX
AC   AAA46500;
XX
DT   04-SEP-2000  (first entry)
XX
DE   cDNA encoding enzyme involved in mannan polysaccharide hydrolysis.
XX
KW   Hydrolysis; polysaccharide; mannan; coffee; endo-beta-mannanase; ds.
XX
OS   Coffea arabica.
XX
FH   Key             Location/Qualifiers
FT   CDS             11..1294
FT                   /*tag= a
FT                   /product= "endo-beta-mannanase"
XX
PN   WO200028046-A1.
XX
PD   18-MAY-2000.
XX
PF   28-OCT-1999;   99WO-EP08314.
XX
PR   11-NOV-1998;   98EP-0203742.
XX
PA   (NEST ) SOC PROD NESTLE SA.
XX
PI   Marraccini P,  Rogers J;
XX
DR   WPI; 2000-399535/34.
DR   P-PSDB; AAY93441.
XX
PT   New DNA encoding endo-beta-mannanase from coffee, used e.g. in
PT   pharmaceutical, cosmetic or food compositions to hydrolyze polymannans
PT   -
XX
PS   Claim 4; Page 29-30; 41pp; French.
XX
CC   The present sequence encodes an endo-beta-mannanase enzyme, which is
CC   involved in the hydrolysis of polysaccharides that consist of molecules
CC   of mannan, either simple or branched, linked together by beta(1-4) bonds.
CC   The mannanase polynucleotide sequence is used for in vivo modification
CC   of the coffee endo-beta-mannanase gene. It is also used to produce
CC   transgenic plant cells (especially coffee cells) which have modified
CC   properties of mannan polysaccharide, and thus altered flavour or
CC   structure. The enzyme is used for modification, degradation or synthesis
CC   of mannan polysaccharides in vitro, particularly to treat coffee beans
CC   to increase the percentage of dry matter extraction, and thus reduce the
CC   quantity of sediment.
XX
SQ   Sequence 1613 BP; 520 A; 305 C; 347 G; 441 T; 0 other;


  Query Match             77.1%;  Score 16.2;  DB 21;  Length 1613;
  Best Local Similarity   85.7%;  Pred. No. 58;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 aagtgaactccctgctacctt 21
          ||||||||||| | || ||||
Db    408 AAGTGAACTCCATACTTCCTT 388


RESULT   7
AAZ09252
ID   AAZ09252 standard; DNA; 32042 BP.
XX
AC   AAZ09252;
XX
DT   25-OCT-1999  (first entry)
XX
DE   Human CARD-4 DNA.
XX
KW   CARD-3; caspase recruitment domain; CARD-4; regulation; detection;
KW   caspase activation; detection; screening; therapy; diagnosis; disease;
KW   apoptotic cell death; Fas/APO-1 receptor complex; TNF receptor complex;
KW   cancer; follicular lymphoma; carcinoma; p53 mutation; viral infection;
KW   hormone-dependent tumour; autoimmune disorder; Alzheimer's disease;
KW   systemic lupus erythematosis; immune-mediated glomerulonephritis; stroke;
KW   Parkinson's disease; amyotrophic lateral sclerosis; retinitis pigmentosa;
KW   spinal muscular dystrophy; cerebellar degeneration; anaemia; drug;
KW   myelodysplastic syndrome; myocardial infarction; cell proliferation;
KW   cell differentiation; cell survival; CARD-4L; CARD-4S; CARD-4Y;
KW   CARD-4Z; human; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   exon            364..685
```

```
FT                      /*tag= a
FT                      /number= 1
FT   CDS                485..31768
FT                      /*tag= b
FT                      /product= "CARD-4"
FT   intron             686..2094
FT                      /*tag= c
FT                      /number= 1
FT   exon               2095..2269
FT                      /*tag= d
FT                      /number= 2
FT   intron             2270..4365
FT                      /*tag= e
FT                      /product= 2
FT   exon               4366..6190
FT                      /*tag= f
FT                      /number= 3
FT   intron             6191..9024
FT                      /*tag= g
FT                      /number= 3
FT   exon               9025..9108
FT                      /*tag= h
FT                      /product= 4
FT   intron             9109..10355
FT                      /*tag= i
FT                      /number= 4
FT   exon               10356..10439
FT                      /*tag= j
FT                      /number= 5
FT   intron             10440..11181
FT                      /*tag= k
FT                      /number= 5
FT   exon               11182..11265
FT                      /*tag= l
FT                      /number= 6
FT   intron             11266..19749
FT                      /*tag= m
FT                      /number= 6
FT   exon               19750..19833
FT                      /*tag= n
FT                      /number= 7
FT   intron             19834..21324
FT                      /*tag= o
FT                      /number= 7
FT   exon               21325..21408
FT                      /*tag= p
FT                      /number= 8
FT   intron             21409..24226
FT                      /*tag= q
FT                      /number= 8
FT   exon               24227..24310
FT                      /*tag= r
FT                      /number= 9
FT   intron             24311..27948
FT                      /*tag= s
FT                      /number= 9
FT   exon               27949..28032
FT                      /*tag= t
FT                      /product= 10
FT   intron             28033..31695
FT                      /*tag= u
FT                      /number= 10
FT   exon               31696..32024
FT                      /*tag= v
FT                      /number= 11
XX
PN   WO9940102-A1.
XX
PD   12-AUG-1999.
XX
PF   05-FEB-1999;   99WO-US02544.
XX
PR   08-DEC-1998;   98US-0207359.
PR   06-FEB-1998;   98US-0019942.
PR   17-JUN-1998;   98US-0099041.
XX
PA   (MILL-) MILLENNIUM PHARM INC.
XX
PI   Bertin J;
XX
DR   WPI; 1999-494269/41.
XX
PT   Novel CARD-3 and CARD-4 genes and polypeptides used or treating
PT   regulation of cellular proliferation and differentiation and cell
PT   survival
XX
PS   Example 13; Fig 18; 181pp; English.
XX
CC   This invention describes the isolation of novel human caspase
CC   recruitment domain, CARD-3 and CARD-4 polynucleotides and proteins and a
CC   partial murine CARD-4L protein and genes. The genes and proteins of
CC   the invention are involved in the regulation of caspase activation.
CC   The caspase recruitment domain (CARD) polynucleotides, polypeptides,
CC   homologues and antibodies can be used in screening assays, detection
CC   assays, predictive medicine and therapeutic and prophylactic methods of
CC   treatment. The methods may be used to diagnose and treat patients which
CC   are suffering from a disorder associated with abnormal level or rate of
CC   apoptotic cell death, abnormal activity of the Fas/APO-1 receptor
CC   complex, abnormal activity of the TNF receptor complex, or abnormal
CC   activity of a caspase. Diseases that may be treated include cancer
CC   (particularly follicular lymphoma, carcinomas associated with mutations
CC   in p53 and hormone-dependent tumours), autoimmune disorders (e.g.
CC   systemic lupus erythematosis, immune-mediated glomerulonephritis), viral
CC   infections, Alzheimer's disease, Parkinson's disease, amyotrophic lateral
CC   sclerosis, retinitis pigmentosa, spinal muscular dystrophy, cerebellar
CC   degeneration, anaemia, myelodysplastic syndrome, myocardial infarction,
CC   and stroke. CARD-3 protein interacts with other cellular proteins, and so
CC   can be used for regulation of cellular proliferation and differentiation
CC   and cell survival. The CARD proteins may also be used to for screen drugs
CC   or compounds which modulate their activity. The CARD-4 gene can express a
CC   long transcript that encodes CARD-4L, a short transcript that encodes
CC   CARD-4S or two CARD-4 splice variants, CARD-4Y and CARD-4Z. This sequence
CC   represents a genomic DNA sequence which encodes the human CARD-4 protein.
XX
SQ   Sequence 32042 BP; 7389 A; 7540 C; 7721 G; 9392 T; 0 other;


  Query Match             77.1%;  Score 16.2;  DB 20;  Length 32042;
  Best Local Similarity   85.7%;  Pred. No. 83;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 aagtgaactccctgctacctt 21
           ||| |||||||||| | ||||
Db   17935 aagggaactccctgatccctt 17955


RESULT   8
AAF30011
ID   AAF30011 standard; cDNA; 32042 BP.
XX
AC   AAF30011;
XX
DT   23-APR-2001  (first entry)
XX
DE   Human CARD-4 gene.
XX
KW   CARD-4; caspase recruitment domain; human; cancer; infection;
KW   autoimmune disease; neurological disease; haematological disease;
KW   immune disease; inflammation; antitumour; antiseptic;
KW   immunomodulator; antiinflammatory; apoptosis; diagnosis;
KW   gene therapy; chromosome 7; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             485..31768
```

```
FT                   /*tag= a
FT                   /note= "contains introns"
FT   exon            364..685
FT                   /*tag= b
FT                   /number= 1
FT   intron          686..2094
FT                   /*tag= c
FT                   /number= 1
FT   exon            2095..2269
FT                   /*tag= d
FT                   /number= 2
FT   intron          2270..4365
FT                   /*tag= e
FT                   /number= 2
FT   exon            4366..6190
FT                   /*tag= f
FT                   /number= 3
FT   intron          6191..9024
FT                   /*tag= g
FT                   /number= 3
FT   exon            9025..9108
FT                   /*tag= h
FT                   /number= 4
FT   intron          9109..10355
FT                   /*tag= i
FT                   /number= 4
FT   exon            10356..10439
FT                   /*tag= j
FT                   /number= 5
FT   intron          10440..11181
FT                   /*tag= k
FT                   /number= 5
FT   exon            11182..11265
FT                   /*tag= l
FT                   /number= 6
FT   intron          11266..19749
FT                   /*tag= m
FT                   /number= 6
FT   exon            19750..19833
FT                   /*tag= n
FT                   /number= 7
FT   intron          19834..21324
FT                   /*tag= o
FT                   /number= 7
FT   exon            21325..21408
FT                   /*tag= p
FT                   /number= 8
FT   intron          21409..24226
FT                   /*tag= q
FT                   /number= 8
FT   exon            24227..24310
FT                   /*tag= r
FT                   /number= 9
FT   intron          24311..27948
FT                   /*tag= s
FT                   /number= 9
FT   exon            27949..28032
FT                   /*tag= t
FT                   /number= 10
FT   intron          28033..31695
FT                   /*tag= u
FT                   /number= 10
FT   exon            31696..32024
FT                   /*tag= v
FT                   /number= 11
XX
PN   WO200100826-A2.
XX
PD   04-JAN-2001.
XX
PF   28-JUN-2000; 2000WO-US17691.
XX
PR   28-JUN-1999;  99US-0340620.
XX
PA   (MILL-) MILLENNIUM PHARM INC.
XX
PI   Bertin J;
XX
DR   WPI; 2001-061973/07.
DR   P-PSDB; AAB20080, AAB20081, AAB20082, AAB20083.
XX
PT   Isolated intracellular proteins predicted to be involved in regulating
PT   caspase activation are used for diagnosis and treatment of e.g. cancer,
PT   viral infections, autoimmune diseases, neurological diseases and
PT   haematological disorders -
XX
PS   Example 13; Fig 18; 208pp; English.
XX
CC   The present sequence is that of the human caspase recruitment
CC   domain 4 (CARD-4) gene on chromosome 7.  CARD-4 exists in at least
CC   4 forms, i.e. the long form CARD-4L (see AAB20080), the short form
CC   CARD-4S (see AAB20081), and splice variants CARD-4Y (see AAB20082)
CC   and CARD-4Z (see AAB20082).  It is an intracellular protein
CC   predicted to be involved in regulating caspase activation.  It
CC   activates the NF-kappaB pathway and enhances caspase 9-mediated
CC   cell death.  Methods of diagnosing and treating patients suffering
CC   from a disorder associated with an abnormal level or rate of apoptotic
CC   cell death, abnormal activity of the Fas/APO-1 receptor complex,
CC   abnormal activity of the tumour necrosis factor receptor complex
CC   or abnormal activity of a caspase involve administering a compound
CC   that modulates the expression or activity of CARD-3, CARD-4, CARD-5
CC   or CARD-6 e.g. a small molecule, antisense nucleic acid, ribozyme
CC   or polypeptide.  Such disorders include cancer, viral infection,
CC   autoimmune disorders, neurological diseases, haematological
CC   disorders, inflammatory disorders and immune disorders.  The CARD-4
CC   gene is useful for genetic information and mapping and identifying
CC   mutations, e.g. mutations in splice donor or acceptor sites.
XX
SQ   Sequence 32042 BP; 7389 A; 7540 C; 7721 G; 9392 T; 0 other;


  Query Match             77.1%;  Score 16.2;  DB 22;  Length 32042;
  Best Local Similarity   85.7%;  Pred. No. 83;
  Matches   18;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 aagtgaactccctgctacctt 21
           ||| |||||||||| | ||||
Db   17935 aagggaactccctgatcccttt 17955


RESULT   9
AAQ97865
ID   AAQ97865 standard; DNA; 957 BP.
XX
AC   AAQ97865;
XX
DT   05-DEC-1995  (first entry)
XX
DE   TROMP1 DNA.
XX
KW   TROMP1; rare outer membrane protein; Spirochaetaceae; immunogen;
KW   vaccine; syphilis; ss.
XX
OS   Treponema pallidum subsp. pallidum.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..957
FT                   /*tag= a
FT   sig_peptide     1..96
FT                   /*tag= b
FT   mat_peptide     97..954
FT                   /*tag= c
XX
PN   WO9518632-A.
XX
PD   13-JUL-1995.
```

XX
PF 06-JAN-1995; 95WO-US00190.
XX
PR 17-AUG-1994; 94US-0292904.
PR 06-JAN-1994; 94US-0178084.
PR 07-JUN-1994; 94US-0255322.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Blanco DR, Champion CI, Lovett MA, Miller JN;
XX
DR WPI; 1995-254907/33.
DR P-PSDB; AAR79722.
XX
PT Rare outer coat membranes of Spirochaetaceae prodn. - by density
PT gradient fractionation, also isolated immunogenic proteins for use in
PT vaccines, derived antibodies and nucleic acid.
XX
PS Disclosure; Page 48-50; 78pp; English.
XX
CC TROMP1 (given in AAR79722) is encoded by DNA isolated using tryptic
CC digest amino acid sequence analysis of EcoRI-digested genomic DNA
CC of T. pallidum subsp. pallidum. It is the precursor of a 31 kDa rare
CC outer membrane protein useful for prepn. of vaccine against syphilis
CC and other treponemal diseases.
XX
SQ Sequence 957 BP; 210 A; 206 C; 303 G; 238 T; 0 other;

Query Match 75.2%; Score 15.8; DB 16; Length 957;
Best Local Similarity 89.5%; Pred. No. 85;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 gtgaactccctgctaccttt 21
|| |||||||||||| |||
Db 749 gtaaactccctgctatctt 767

RESULT 10
AAT76580
ID AAT76580 standard; DNA; 957 BP.
XX
AC AAT76580;
XX
DT 13-JAN-1998 (first entry)
XX
DE Treponema pallidum rare outer membrane protein (TROMP-1) DNA.
XX
KW Rare; outer membrane protein; Treponema pallidum; Spirochaetales;
KW Immune response; Syphilis; TROMP-1; precursor; ss.
XX
OS Treponema pallidum.
XX
FH Key Location/Qualifiers
FT CDS 1..957
FT /*tag= a
FT /product= immature_TROMP1_protein
XX
PN WO9727215-A1.
XX
PD 31-JUL-1997.
XX
PF 23-JAN-1997; 97WO-US01302.
XX
PR 23-JAN-1996; 96US-0599480.
XX
PA (SLOK) SLOAN KETTERING INST CANCER RES.
PA (REGC) UNIV CALIFORNIA.
XX
PI Blanco DR, Lovitt MA, Miller JN, Tempst PJ;
XX
DR WPI; 1997-393614/36.
DR P-PSDB; AAW22134.
XX
PT New isolated Treponema pallidum outer membrane protein - used in the
PT detection and induction of immune response in an animal to
PT pathogenic Spirochaetales
XX
PS Disclosure; Pages 51-53; 77pp; English.
XX
CC This 957 bp sequence (TROMP1 gene) encodes a precursor rare outer
CC membrane protein (OMP) of species Treponema pallidum, subspecies
CC pallidum. The TROMP1 gene was cloned in a procedure where mixed
CC oligonucleotides (31-A,31-C) hybridized to a .EcoR1 restriction
CC fragment by Southern blot analysis of T. pallidum genomic DNA. DNA
CC fragments were excised from the agarose gel, purified and ligated
CC into the lambda ZAP II vector and probed with 31-A and 31-C. The
CC phage clones were converted to the pBluescript SK(-) recombinant
CC plasmid by in vivo excision. The recombinant expression of these
CC rare OMP's can be used for diagnostic tests to detect syphilis and
CC for development of host immunity during syphilis.
XX
SQ Sequence 957 BP; 210 A; 206 C; 303 G; 238 T; 0 other;

Query Match 75.2%; Score 15.8; DB 18; Length 957;
Best Local Similarity 89.5%; Pred. No. 85;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 gtgaactccctgctaccttt 21
|| |||||||||||| |||
Db 749 gtaaactccctgctatctt 767

RESULT 11
AAA67251/c
ID AAA67251 standard; DNA; 537 BP.
XX
AC AAA67251;
XX
DT 31-OCT-2000 (first entry)
XX
DE Eucalyptus grandis beta-amylase DNA sequence SEQ ID NO:252.
XX
KW Eucalyptus grandis; pinus radiata; Monterey pine; modification;
KW plant cell wall; polysaccharide; polysaccharide biosynthetic pathway;
KW transgenic plant; ds.
XX
OS Eucalyptus grandis.
XX
PN WO200022092-A2.
XX
PD 20-APR-2000.
XX
PF 08-OCT-1999; 99WO-NZ00169.
XX
PR 13-OCT-1998; 98US-0170862.
PR 11-AUG-1999; 99US-0148426.
XX
PA (GENE-) GENESIS RES & DEV CORP LTD.
PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
PI Bloksberg LN;
XX
DR WPI; 2000-339328/29.
XX
PT New genes encoding proteins involved in a plant polysaccharide
PT biosynthetic pathway, useful for modulating or altering the
PT polysaccharide content, composition or structure of the plant -
XX
PS Claim 1; Page 140; 301pp; English.
XX
CC The present invention describes isolated polynucleotides (PN) comprising
CC a sequence selected from one of 835 nucleotide sequences given in
CC AAA67073 to AAA67907, their (reverse) complements, sequences producing
CC an Expectation (E) value of 0.01 or less compared to the 835 sequences,

CC sequences at least 50% identical to them, 200, 100, 40 or 20-mers of the
CC 835 sequences or sequences that are degenerately equivalent or allelic
CC to the 835 sequences. The polynucleotides are used to modify the
CC activity of a polypeptide involved in a polysaccharide biosynthetic
CC pathway in the plant. They are especially used to modulate or alter the
CC polysaccharide content, composition or structure of the plant. AAB16268
CC to AAB16340 are proteins encoded by some of the polynucleotide sequence
CC given in the present invention.
XX
SQ Sequence 537 BP; 155 A; 119 C; 136 G; 127 T; 0 other;

Query Match 72.4%; Score 15.2; DB 21; Length 537;
Best Local Similarity 85.0%; Pred. No. 1.6e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 agtgaactccctgctacctt 21
||||||| | ||||| |||||
Db 472 AGTGAAATTCCTGCAACCTT 453

RESULT 12
AAV04426
ID AAV04426 standard; cDNA; 640 BP.
XX
AC AAV04426;
XX
DT 27-APR-1998 (first entry)
XX
DE Human calpain small subunit cDNA fragment.
XX
KW Calpain; human; leukocyte; calcium dependent cysteine protease;
KW screening; activator; inhibitor; treatment; prevention; cancer;
KW cerebral apoplexy; cerebral infarction; subarachnoid haemorrhage;
KW Alzheimer's disease; myodystrophy; cataracts; collagen disease;
KW ischaemic heart disease; atherosclerosis; arthritis;
KW small subunit; ds.
XX
OS Homo sapiens.
XX
PN EP799892-A2.
XX
PD 08-OCT-1997.
XX
PF 03-APR-1997; 97EP-0105508.
XX
PR 05-APR-1996; 96JP-0083649.
XX
PA (TAKE) TAKEDA CHEM IND LTD.
XX
PI Kawamoto T, Nishi K, Shintani Y;
XX
DR WPI; 1997-482674/45.
XX
PT Human calpain protein and related DNA - useful for drug screening
PT and treating cancer, stroke, etc.
XX
PS Example 2; Page 38; 43pp; English.
XX
CC The present sequence is a calpain small subunit cDNA fragment.
CC Calpain is a human leukocyte derived calcium dependent cysteine
CC protease. Calpain can be used to screen for compounds that activate
CC or inhibit its proteolytic activity. Calpain DNA can be used to
CC treat or prevent cancer, cerebral apoplexy, cerebral infarction,
CC subarachnoid haemorrhage, Alzheimer's disease, myodystrophy,
CC cataracts, ischaemic heart disease, atherosclerosis, arthritis
CC or collagen disease.
XX
SQ Sequence 640 BP; 139 A; 163 C; 223 G; 115 T; 0 other;

Query Match 72.4%; Score 15.2; DB 18; Length 640;
Best Local Similarity 85.0%; Pred. No. 1.6e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 agtgaactccctgctacctt 21
||||||||||| | | ||||
Db 574 agtgaactcccaggtgcctt 593

RESULT 13
AAV04420
ID AAV04420 standard; cDNA; 804 BP.
XX
AC AAV04420;
XX
DT 27-APR-1998 (first entry)
XX
DE Human calpain small subunit cDNA.
XX
KW Calpain; human; leukocyte; calcium dependent cysteine protease;
KW screening; activator; inhibitor; treatment; prevention; cancer;
KW cerebral apoplexy; cerebral infarction; subarachnoid haemorrhage;
KW Alzheimer's disease; myodystrophy; cataracts; collagen disease;
KW ischaemic heart disease; atherosclerosis; arthritis;
KW small subunit; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1..2109
FT /*tag= a
FT /note= "stop codon not given"
XX
PN EP799892-A2.
XX
PD 08-OCT-1997.
XX
PF 03-APR-1997; 97EP-0105508.
XX
PR 05-APR-1996; 96JP-0083649.
XX
PA (TAKE) TAKEDA CHEM IND LTD.
XX
PI Kawamoto T, Nishi K, Shintani Y;
XX
DR WPI; 1997-482674/45.
DR P-PSDB; AAW41640.
XX
PT Human calpain protein and related DNA - useful for drug screening
PT and treating cancer, stroke, etc.
XX
PS Disclosure; Page 35; 43pp; English.
XX
CC The present sequence encodes calpain small subunit, a human
CC leukocyte derived calcium dependent cysteine protease. Calpain can
CC be used to screen for compounds that activate or inhibit its
CC proteolytic activity. Calpain DNA can be used to treat or prevent
CC cancer, cerebral apoplexy, cerebral infarction, subarachnoid
CC haemorrhage, Alzheimer's disease, myodystrophy, cataracts,
CC ischaemic heart disease, atherosclerosis, arthritis or collagen
CC disease.
XX
SQ Sequence 804 BP; 181 A; 203 C; 264 G; 156 T; 0 other;

Query Match 72.4%; Score 15.2; DB 18; Length 804;
Best Local Similarity 85.0%; Pred. No. 1.7e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 agtgaactccctgctacctt 21
||||||||||| | | ||||
Db 574 agtgaactcccaggtgcctt 593

RESULT 14

AAC77765
ID AAC77765 standard; cDNA; 1242 BP.
XX
AC AAC77765;
XX
DT 08-FEB-2001 (first entry)
XX
DE Human cancer associated gene sequence SEQ ID NO:159.
XX
KW Human; cancer associated gene; cancer antigen; detection; cancer;
KW diagnosis; cytostatic; proliferative; vulnerary; immunomodulator;
KW antidiabetic; antiasthmatic; antirheumatic; antiarthritic; antiviral;
KW antiinflammatory; antithyroid; antiallergic; antibacterial; cardiant;
KW dermatological; neuroprotective; thrombolytic; coagulant; nootropic;
KW vasotropic; antipsoriatic; antiangiogenic; gene therapy; inflammation;
KW immune disorder; haematopoietic cell disorder; autoimmune disorder;
KW allergic reaction; graft versus host disease; organ rejection;
KW haemostatic; thrombolytic; cardiovascular disorder; infection;
KW neurological disease; drug screening; ss.
XX
OS Homo sapiens.
XX
PN WO200055350-A1.
XX
PD 21-SEP-2000.
XX
PF 08-MAR-2000; 2000WO-US05882.
XX
PR 12-MAR-1999; 99US-0124270.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM;
XX
DR WPI; 2000-587533/55.
DR P-PSDB; AAB43556.
XX
PT Novel isolated nucleic acids comprising sequences encoding peptides
PT useful for treating or diagnosing e.g. cancer -
XX
PS Claim 1; Page 737-738; 2352pp; English.
XX
CC AAC77607 to AAC78448 encode the human cancer associated proteins given
CC in AAB43398 to AAB44239. The proteins can have activities based on the
CC tissues and cells the genes are expressed in. Example of activities
CC include: cytostatic; proliferative; vulnerary; immunomodulator;
CC antidiabetic; antiasthmatic; antirheumatic; antiarthritic;
CC antiinflammatory; antithyroid; antiallergic; antibacterial; antiviral;
CC dermatological; neuroprotective; cardiant; thrombolytic; coagulant;
CC nootropic; vasotropic; antipsoriatic and antiangiogenic. The
CC polynucleotides and polypeptides can be used for preventing, treating or
CC ameliorating medical conditions and diagnosing pathological conditions.
CC Polynucleotides, polypeptides, antibodies, agonists and antagonists from
CC the present invention may be used to treat immune disorders by activating
CC or inhibiting the proliferation, differentiation or mobilisation of
CC immune cells, to treat disorders of haematopoietic cells, autoimmune
CC disorders, allergic reactions, graft versus host disease and organ
CC rejection, modulate haemostatic or thrombolytic activity, modulate
CC inflammation, cancers, cardiovascular disorders, neurological disease and
CC bacterial or viral infections. The peptides, nucleotides, antibodies,
CC agonists and antagonists may be also be used in drug screens. AAC78449 to
CC AAC78457 and AAB44240 represent sequences used in the exemplification of
CC the present invention.
XX
SQ Sequence 1242 BP; 333 A; 357 C; 292 G; 258 T; 2 other;

Query Match 72.4%; Score 15.2; DB 21; Length 1242;
Best Local Similarity 85.0%; Pred. No. 1.8e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 agtgaactccctgctacctt 21
||||||||||| | | ||||
Db 414 agtgaactcccaggtgcctt 433

RESULT 15
AAZ98118/c
ID AAZ98118 standard; cDNA; 1741 BP.
XX
AC AAZ98118;
XX
DT 11-MAY-2000 (first entry)
XX
DE Human signal peptide containing protein HSPP-10 cDNA SEQ ID NO:144.
XX
KW Human; signal peptide-containing protein; HSPP; diagnosis; cancer;
KW inflammation; cardiovascular disease; anticancer; anti-inflammatory;
KW antimicrobial; nootropic; neuroprotective; cardiovascular; hepatotropic;
KW antiasthmatic; gene therapy; cell proliferation; neurological disorder;
KW reproductive disorder; developmental disorder; arteriosclerosis;
KW cirrhosis; psoriasis; acquired immune deficiency syndrome; anaemia;
KW asthma; Crohn's disease; infection; Alzheimer's disease; schizophrenia;
KW Parkinson's disease; Huntington's diseases; ovulatory defect;
KW muscular dystrophy; ss.
XX
OS Homo sapiens.
XX
PN WO200000610-A2.
XX
PD 06-JAN-2000.
XX
PF 25-JUN-1999; 99WO-US14484.
XX
PR 26-JUN-1998; 98US-0090762.
PR 31-JUL-1998; 98US-0094983.
PR 01-OCT-1998; 98US-0102686.
PR 11-DEC-1998; 98US-0112129.
XX
PA (INCY-) INCYTE PHARM INC.
XX
PI Lal P, Tang YT, Gorgone GA, Corley NC, Guegler KJ, Baughn MR;
PI Akerblom IE, Au-Young J, Yue H, Patterson C, Reddy R, Hillman JL;
PI Bandman O;
XX
DR WPI; 2000-160673/14.
DR P-PSDB; AAY87233.
XX
PT New human signal peptide-containing proteins useful in treatment,
PT prevention and diagnosis of e.g. cancer, inflammation and
PT cardiovascular disease -
XX
PS Claim 9; Page 257-258; 327pp; English.
XX
CC AAZ98109 to AAZ98242 encode AAY87224 to AAY87357 which represent the
CC human signal peptide-containing proteins HSPP-1 to HSPP-134. HSPPs have
CC anticancer, anti-inflammatory, antimicrobial, nootropic, hepatotropic,
CC neuroprotective, cardiovascular and antiasthmatic activities, and can
CC be used in gene therapy. HSPPs can be used to treat or prevent disorders
CC associated with decreased activity or function of HSPP. Antagonists of
CC HSPP are used to treat or prevent disorders associated with increased
CC activity or function of HSPP. Such diseases include cell proliferation
CC (including cancer), inflammation, cardiovascular, neurological,
CC reproductive or developmental disorders, (e.g. arteriosclerosis,
CC cirrhosis, psoriasis, acquired immune deficiency syndrome, anaemia,
CC asthma, Crohn's disease, microbial or other infections, congestive or
CC ischaemic heart disease, Alzheimer's, Parkinson's or Huntington's
CC diseases, schizophrenia, ovulatory defects, muscular dystrophy). HSPP
CC nucleic acids can be used for the recombinant production of HSPP, for
CC detecting HSPP in standard hybridisation and amplification assays (for
CC diagnosis and monitoring), in gene therapy, as antisense,
CC triplex-forming or ribozyme therapeutics, for detecting related sequences
CC or genetic variations, and for chromosomal mapping. HSPP are also used to
CC raise specific antibodies (Ab) and to screen for agonists and
CC antagonists (potential therapeutic agents). Ab are used to diagnose, or
CC monitor, HSPP-related diseases (in usual immunoassays), as therapeutic

CC antagonists, in competitive drug screens, and for purification of HSPP
CC from natural sources.
XX
SQ Sequence 1741 BP; 533 A; 358 C; 364 G; 486 T; 0 other;

Query Match 72.4%; Score 15.2; DB 21; Length 1741;
Best Local Similarity 85.0%; Pred. No. 1.8e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
Qy       1 aagtgaactccctgctacct 20
           ||||||| |  ||||||||||
Db    1411 AAGTGAAGTGTCTGCTACCT 1392
```

Search completed: July 25, 2001, 05:23:03
Job time: 4676 sec

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: July 25, 2001, 02:45:17 ; Search time 117.39 Seconds
(without alignments)
33.141 Million cell updates/sec

Title: US-09-142-095-1
Perfect score: 21
Sequence: 1 aagtgaactccctgctacctt 21

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 317530 seqs, 92630169 residues

Total number of hits satisfying chosen parameters: 635060

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	15.8	75.2	957	1	US-08-599-480-1	Sequence 1, Appli
	2	15.8	75.2	957	1	US-08-842-199-1	Sequence 1, Appli
	3	15.8	75.2	957	5	PCT-US95-00190-1	Sequence 1, Appli
	4	15.2	72.4	640	2	US-08-835-099A-16	Sequence 16, Appl
	5	15.2	72.4	640	3	US-09-157-349-16	Sequence 16, Appl
	6	15.2	72.4	804	2	US-08-835-099A-10	Sequence 10, Appl
	7	15.2	72.4	804	3	US-09-157-349-10	Sequence 10, Appl
c	8	15.2	72.4	246240	2	US-08-724-394A-20	Sequence 20, Appl
c	9	15.2	72.4	246240	2	US-08-724-394A-21	Sequence 21, Appl
c	10	15.2	72.4	246240	2	US-08-724-394A-22	Sequence 22, Appl
	11	14.8	70.5	1723	1	US-08-417-492-1	Sequence 1, Appli
	12	14.8	70.5	2161	2	US-09-001-826-4	Sequence 4, Appli
	13	14.8	70.5	2161	2	US-09-001-826-25	Sequence 25, Appl
	14	14.8	70.5	2652	2	US-08-366-547-1	Sequence 1, Appli
c	15	14.8	70.5	2977	1	US-08-276-151-1	Sequence 1, Appli
c	16	14.8	70.5	2977	1	US-08-306-691B-17	Sequence 17, Appl
c	17	14.8	70.5	2977	2	US-08-756-806A-64	Sequence 64, Appl
c	18	14.8	70.5	2977	3	US-08-328-239A-4	Sequence 4, Appli
c	19	14.8	70.5	2977	3	US-09-143-214-64	Sequence 64, Appl
c	20	14.8	70.5	2977	4	US-09-209-668-12	Sequence 12, Appl
c	21	14.8	70.5	2977	5	PCT-US95-13661-4	Sequence 4, Appli
c	22	14.6	69.5	394	1	US-08-488-144-18	Sequence 18, Appl
	23	14.6	69.5	1168	4	US-08-930-285-15	Sequence 15, Appl
	24	14.6	69.5	4345	2	US-08-244-537-1	Sequence 1, Appli
	25	14.6	69.5	7470	3	US-08-417-089-5	Sequence 5, Appli
	26	14.6	69.5	7470	4	US-08-695-651-5	Sequence 5, Appli
	27	14.6	69.5	7470	4	US-08-930-285-5	Sequence 5, Appli
	28	14.4	68.6	1273	4	US-08-725-758A-3	Sequence 3, Appli
	29	14.4	68.6	1373	4	US-08-725-758A-1	Sequence 1, Appli
	30	14.4	68.6	4310	4	US-09-008-172-1	Sequence 1, Appli
	31	14.2	67.6	61	4	US-08-687-421-355	Sequence 355, App
	32	14.2	67.6	1118	1	US-08-418-032-1	Sequence 1, Appli
	33	14.2	67.6	1554	1	US-08-469-486-1	Sequence 1, Appli
	34	14.2	67.6	1554	2	US-08-469-658-1	Sequence 1, Appli
c	35	14.2	67.6	3504	1	US-08-620-717A-8	Sequence 8, Appli
c	36	14.2	67.6	5541	1	US-08-920-812-20	Sequence 20, Appl
c	37	14.2	67.6	5541	1	US-08-920-827-20	Sequence 20, Appl
c	38	14.2	67.6	5541	1	US-08-921-177-20	Sequence 20, Appl
c	39	14.2	67.6	5541	1	US-08-362-577C-20	Sequence 20, Appl
c	40	14.2	67.6	5541	2	US-08-920-828-20	Sequence 20, Appl
	41	14.2	67.6	20303	1	US-08-370-975B-6	Sequence 6, Appli
	42	14.2	67.6	26764	1	US-08-370-975B-1	Sequence 1, Appli
	43	14	66.7	1441	1	US-08-136-277-18	Sequence 18, Appl
	44	14	66.7	1441	2	US-08-479-403-18	Sequence 18, Appl
	45	14	66.7	1441	3	US-08-835-734-18	Sequence 18, Appl

ALIGNMENTS

```
RESULT 1
US-08-599-480-1
; Sequence 1, Application US/08599480
; Patent No. 5753459
;  GENERAL INFORMATION:
;    APPLICANT:  Blanco, David R.
;    APPLICANT:  Miller, James N.
;    APPLICANT:  Lovett, Michael A.
;    APPLICANT:  Champion, Cheryl I.
;    APPLICANT:  Tempst, Paul J.
;    TITLE OF INVENTION:  NUCLEOTIDE AND AMINO ACID SEQUENCES OF A
;    TITLE OF INVENTION:  T. Pallidum RARE OUTER MEMBRANE PROTEIN
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson P.C.
;      STREET:  4225 Executive Square, Suite 1400
;      CITY:  La Jolla
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  92037
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/599,480
;      FILING DATE:  23-JAN-1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Learn, June M.
;      REGISTRATION NUMBER:  31,238
;      REFERENCE/DOCKET NUMBER:  07419/018001 (CIP of 016001)
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 678-5070
;      TELEFAX:  (619) 678-5099
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  957 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    IMMEDIATE SOURCE:
;      CLONE:  TROMP1
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..954
US-08-599-480-1
```

```
  Query Match           75.2%;  Score 15.8;  DB 1;  Length 957;
  Best Local Similarity  89.5%;  Pred. No. 19;
  Matches   17;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      3 gtgaactccctgctacctt 21
          || |||||||||||| |||
Db    749 GTAAACTCCCTGCTATCTT 767


RESULT   2
US-08-842-199-1
; Sequence 1, Application US/08842199
; Patent No. 5821085
;  GENERAL INFORMATION:
;    APPLICANT:  Blanco, David R.
;    APPLICANT:  Miller, James N.
;    APPLICANT:  Lovett, Michael A.
;    APPLICANT:  Champion, Cheryl I.
;    TITLE OF INVENTION:  NUCLEOTIDE AND AMINO ACID SEQUENCES OF A
;    TITLE OF INVENTION:  T. Pallidum RARE OUTER MEMBRANE PROTEIN
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Spensley Horn Jubas & Lubitz
;      STREET:  1880 Century Park East, Suite 500
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  90067
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/842,199
;      FILING DATE:  23-APR-1997
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/292,904
;      FILING DATE:  17-AUG-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bostich, June M.
;      REGISTRATION NUMBER:  31,238
;      REFERENCE/DOCKET NUMBER:  PD-3744
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 455-5100
;      TELEFAX:  (519) 455-5110
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  957 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    IMMEDIATE SOURCE:
;      CLONE:  TROMP1
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..954
US-08-842-199-1


  Query Match           75.2%;  Score 15.8;  DB 1;  Length 957;
  Best Local Similarity  89.5%;  Pred. No. 19;
  Matches   17;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      3 gtgaactccctgctacctt 21
          || |||||||||||| |||
Db    749 GTAAACTCCCTGCTATCTT 767


RESULT   3
PCT-US95-00190-1
; Sequence 1, Application PC/TUS9500190
;  GENERAL INFORMATION:
;    APPLICANT:  THE REGENTS OF THE UNIVERSITY OF CALIFORNIA
;    TITLE OF INVENTION:  NUCLEOTIDE AND AMINO ACID SEQUENCES OF A
;    TITLE OF INVENTION:  T. Pallidum RARE OUTER MEMBRANE PROTEIN
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Spensley Horn Jubas & Lubitz
;      STREET:  1880 Century Park East, Suite 500
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  90067
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/00190
;      FILING DATE:  06-JAN-1995
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bostich, June M.
;      REGISTRATION NUMBER:  31,238
;      REFERENCE/DOCKET NUMBER:  PD-3744
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 455-5100
;      TELEFAX:  (619) 455-5110
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  957 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    IMMEDIATE SOURCE:
;      CLONE:  TROMP1
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..954
PCT-US95-00190-1


  Query Match           75.2%;  Score 15.8;  DB 5;  Length 957;
  Best Local Similarity  89.5%;  Pred. No. 19;
  Matches   17;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      3 gtgaactccctgctacctt 21
          || |||||||||||| |||
Db    749 GTAAACTCCCTGCTATCTT 767


RESULT   4
US-08-835-099A-16
; Sequence 16, Application US/08835099A
; Patent No. 5874277
;  GENERAL INFORMATION:
;    APPLICANT:  SHINTANI, Yasushi
;    APPLICANT:  NISHI, Kaznori
;    APPLICANT:  KAWAMOTO, Tomohiro
;    TITLE OF INVENTION:  NOVEL PROTEINS, THEIR PRODUCTION
;    TITLE OF INVENTION:  AND USE
;    NUMBER OF SEQUENCES:  18
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DIKE, BRONSTEIN, ROBERTS & CUSMAN, LLP
;      STREET:  130 Water Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
```

```
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/835,099A
;      FILING DATE:  04-APR-1997
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  083649/1996
;      FILING DATE:  05-APR-1996
;      APPLICATION NUMBER:  97105508.2
;      FILING DATE:  03-APR-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Resnick, David S
;      REGISTRATION NUMBER:  34,235
;      REFERENCE/DOCKET NUMBER:  47342
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-523-3400
;      TELEFAX:  617-523-6440
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  16:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  640 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-835-099A-16


  Query Match             72.4%;  Score 15.2;  DB 2;  Length 640;
  Best Local Similarity   85.0%;  Pred. No. 36;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 agtgaactccctgctacctt 21
           |||||||||||  | |  ||||
Db     574 AGTGAACTCCCAGGTGCCTT 593


RESULT   5
US-09-157-349-16
; Sequence 16, Application US/09157349
; Patent No. 6068990
;  GENERAL INFORMATION:
;    APPLICANT:  SHINTANI, Yasushi
;    APPLICANT:  NISHI, Kaznori
;    APPLICANT:  KAWAMOTO, Tomohiro
;    TITLE OF INVENTION:  NOVEL PROTEINS, THEIR PRODUCTION
;    TITLE OF INVENTION:  AND USE
;    NUMBER OF SEQUENCES:  18
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DIKE, BRONSTEIN, ROBERTS & CUSMAN, LLP
;      STREET:  130 Water Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/157,349
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/835,099
;      FILING DATE:
;      APPLICATION NUMBER:  97105508.2
;      FILING DATE:  03-APR-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Resnick, David S
;      REGISTRATION NUMBER:  34,235
;      REFERENCE/DOCKET NUMBER:  47342
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-523-3400
;      TELEFAX:  617-523-6440
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  16:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  640 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-09-157-349-16


  Query Match             72.4%;  Score 15.2;  DB 3;  Length 640;
  Best Local Similarity   85.0%;  Pred. No. 36;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 agtgaactccctgctacctt 21
           |||||||||||  | |  ||||
Db     574 AGTGAACTCCCAGGTGCCTT 593


RESULT   6
US-08-835-099A-10
; Sequence 10, Application US/08835099A
; Patent No. 5874277
;  GENERAL INFORMATION:
;    APPLICANT:  SHINTANI, Yasushi
;    APPLICANT:  NISHI, Kaznori
;    APPLICANT:  KAWAMOTO, Tomohiro
;    TITLE OF INVENTION:  NOVEL PROTEINS, THEIR PRODUCTION
;    TITLE OF INVENTION:  AND USE
;    NUMBER OF SEQUENCES:  18
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DIKE, BRONSTEIN, ROBERTS & CUSMAN, LLP
;      STREET:  130 Water Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/835,099A
;      FILING DATE:  04-APR-1997
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  083649/1996
;      FILING DATE:  05-APR-1996
;      APPLICATION NUMBER:  97105508.2
;      FILING DATE:  03-APR-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Resnick, David S
;      REGISTRATION NUMBER:  34,235
;      REFERENCE/DOCKET NUMBER:  47342
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-523-3400
;      TELEFAX:  617-523-6440
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  804 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
```

US-08-835-099A-10

```
  Query Match           72.4%;  Score 15.2;  DB 2;  Length 804;
  Best Local Similarity  85.0%;  Pred. No. 37;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 agtgaactccctgctacctt 21
           ||||||||||| | | ||||
Db     574 AGTGAACTCCCAGGTGCCTT 593
```

```
RESULT   7
US-09-157-349-10
; Sequence 10, Application US/09157349
; Patent No. 6068990
;  GENERAL INFORMATION:
;    APPLICANT:  SHINTANI, Yasushi
;    APPLICANT:  NISHI, Kaznori
;    APPLICANT:  KAWAMOTO, Tomohiro
;    TITLE OF INVENTION:  NOVEL PROTEINS, THEIR PRODUCTION
;    TITLE OF INVENTION:  AND USE
;    NUMBER OF SEQUENCES:  18
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DIKE, BRONSTEIN, ROBERTS & CUSMAN, LLP
;      STREET:  130 Water Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/157,349
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/835,099
;      FILING DATE:
;      APPLICATION NUMBER:  97105508.2
;      FILING DATE:  03-APR-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Resnick, David S
;      REGISTRATION NUMBER:  34,235
;      REFERENCE/DOCKET NUMBER:  47342
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-523-3400
;      TELEFAX:  617-523-6440
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  804 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-09-157-349-10
```

```
  Query Match           72.4%;  Score 15.2;  DB 3;  Length 804;
  Best Local Similarity  85.0%;  Pred. No. 37;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 agtgaactccctgctacctt 21
           ||||||||||| | | ||||
Db     574 AGTGAACTCCCAGGTGCCTT 593
```

```
RESULT   8
US-08-724-394A-20/c
; Sequence 20, Application US/08724394A
; Patent No. 5872237
;  GENERAL INFORMATION:
;    APPLICANT:  Feder, John N.
;    APPLICANT:  Kronmal, Gregory S.
;    APPLICANT:  Lauer, Peter M.
;    APPLICANT:  Ruddy, David A.
;    APPLICANT:  Thomas, Winston
;    APPLICANT:  Tsuchihashi, Zenta
;    APPLICANT:  Wolff, Roger K.
;    TITLE OF INVENTION:  Megabase Transcript Map:  No. 5872237el
;    TITLE OF INVENTION:  Sequences and Antibodies Thereto
;    NUMBER OF SEQUENCES:  31
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  TOWNSEND and TOWNSEND and CREW LLP
;      STREET:  Two Embarcadero Center, 8th Floor
;      CITY:  San Francisco
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94111-3834
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/724,394A
;      FILING DATE:  01-OCT-1996
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fitts, Renee A.
;      REGISTRATION NUMBER:  35,136
;      REFERENCE/DOCKET NUMBER:  017957-000100
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-576-0200
;      TELEFAX:  415-576-0300
;  INFORMATION FOR SEQ ID NO:  20:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  246240 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  not relevant
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1..246240
;      OTHER INFORMATION:  /note= "HLA-H.CONTIG"
US-08-724-394A-20
```

```
  Query Match           72.4%;  Score 15.2;  DB 2;  Length 246240;
  Best Local Similarity  85.0%;  Pred. No. 86;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 aagtgaactccctgctacct 20
           | |||||||| |||| ||||
Db 139464 ATGTGAACTCACTGCAACCT 139445
```

```
RESULT   9
US-08-724-394A-21/c
; Sequence 21, Application US/08724394A
; Patent No. 5872237
;  GENERAL INFORMATION:
;    APPLICANT:  Feder, John N.
;    APPLICANT:  Kronmal, Gregory S.
;    APPLICANT:  Lauer, Peter M.
;    APPLICANT:  Ruddy, David A.
;    APPLICANT:  Thomas, Winston
;    APPLICANT:  Tsuchihashi, Zenta
;    APPLICANT:  Wolff, Roger K.
;    TITLE OF INVENTION:  Megabase Transcript Map:  No. 5872237el
;    TITLE OF INVENTION:  Sequences and Antibodies Thereto
```

```
;    NUMBER OF SEQUENCES:  31
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  TOWNSEND and TOWNSEND and CREW LLP
;      STREET:  Two Embarcadero Center, 8th Floor
;      CITY:  San Francisco
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94111-3834
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/724,394A
;      FILING DATE:  01-OCT-1996
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fitts, Renee A.
;      REGISTRATION NUMBER:  35,136
;      REFERENCE/DOCKET NUMBER:  017957-000100
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-576-0200
;      TELEFAX:  415-576-0300
;  INFORMATION FOR SEQ ID NO:  21:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  246240 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  not relevant
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1..246240
;      OTHER INFORMATION:  /note= "HLA-H.CONTIG"
US-08-724-394A-21


  Query Match             72.4%;  Score 15.2;  DB 2;  Length 246240;
  Best Local Similarity   85.0%;  Pred. No. 86;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 aagtgaactccctgctacct 20
          | ||||||| |||| ||||
Db 139464 ATGTGAACTCACTGCAACCT 139445


RESULT  10
US-08-724-394A-22/c
; Sequence 22, Application US/08724394A
; Patent No. 5872237
;  GENERAL INFORMATION:
;    APPLICANT:  Feder, John N.
;    APPLICANT:  Kronmal, Gregory S.
;    APPLICANT:  Lauer, Peter M.
;    APPLICANT:  Ruddy, David A.
;    APPLICANT:  Thomas, Winston
;    APPLICANT:  Tsuchihashi, Zenta
;    APPLICANT:  Wolff, Roger K.
;    TITLE OF INVENTION:  Megabase Transcript Map:  No. 5872237el
;    TITLE OF INVENTION:  Sequences and Antibodies Thereto
;    NUMBER OF SEQUENCES:  31
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  TOWNSEND and TOWNSEND and CREW LLP
;      STREET:  Two Embarcadero Center, 8th Floor
;      CITY:  San Francisco
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94111-3834
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/724,394A
;      FILING DATE:  01-OCT-1996
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fitts, Renee A.
;      REGISTRATION NUMBER:  35,136
;      REFERENCE/DOCKET NUMBER:  017957-000100
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-576-0200
;      TELEFAX:  415-576-0300
;  INFORMATION FOR SEQ ID NO:  22:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  246240 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  not relevant
;      TOPOLOGY:  not relevant
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1..246240
;      OTHER INFORMATION:  /note= "HLA-H.CONTIG"
US-08-724-394A-22


  Query Match             72.4%;  Score 15.2;  DB 2;  Length 246240;
  Best Local Similarity   85.0%;  Pred. No. 86;
  Matches   17;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 aagtgaactccctgctacct 20
          | ||||||| |||| ||||
Db 139464 ATGTGAACTCACTGCAACCT 139445


RESULT  11
US-08-417-492-1
; Sequence 1, Application US/08417492
; Patent No. 5750872
;  GENERAL INFORMATION:
;    APPLICANT:  Bennett, Alan B
;    APPLICANT:  Brummell, David A
;    APPLICANT:  Grantz, Alexander A
;    TITLE OF INVENTION:  Nucleic Acids Encoding Ascorbate Free
;    TITLE OF INVENTION:  Radical Reductase and Their Uses
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Townsend and Townsend and Crew
;      STREET:  One Market Plaza, Steuart Street Tower
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94105-1492
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/417,492
;      FILING DATE:  05-APR-1995
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bastian, Kevin L
;      REGISTRATION NUMBER:  34,774
;      REFERENCE/DOCKET NUMBER:  2307E-586US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-543-9600
;      TELEFAX:  415-543-5043
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1723 base pairs
;      TYPE:  nucleic acid
```

```
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     HYPOTHETICAL:  NO
;     FEATURE:
;       NAME/KEY:  exon
;       LOCATION:  49..643
;       OTHER INFORMATION:  /number= 1
;     FEATURE:
;       NAME/KEY:  intron
;       LOCATION:  644..780
;       OTHER INFORMATION:  /number= 1
;     FEATURE:
;       NAME/KEY:  exon
;       LOCATION:  781..1484
;       OTHER INFORMATION:  /number= 2
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  join(49..643, 781..1487)
US-08-417-492-1


  Query Match             70.5%;  Score 14.8;  DB 1;  Length 1723;
  Best Local Similarity   88.9%;  Pred. No. 68;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       4 tgaactccctgctacctt 21
           | ||| ||||||||||||
Db    1424 TCAACCCCCTGCTACCTT 1441


RESULT  12
US-09-001-826-4
; Sequence 4, Application US/09001826A
; Patent No. 5965727
; GENERAL INFORMATION:
;  APPLICANT: SONG, HEE-SOOK
;  APPLICANT: BROTHERTON, JEFFREY E.
;  APPLICANT: WIDHOLM, JACK M.
;  TITLE OF INVENTION: SELECTABLE MARKER AND PROMOTER FOR PLANT TISSUE CULTURE
;  TITLE OF INVENTION: TRANSFORMATION
;  FILE REFERENCE: UI001.C1
;  CURRENT APPLICATION NUMBER: US/09/001,826A
;  CURRENT FILING DATE: 1997-12-31
;  EARLIER APPLICATION NUMBER: 08/937,739;  60/025,140
;  EARLIER FILING DATE: 1997-07-25; 1996-07-26
;  NUMBER OF SEQ ID NOS: 31
;  SOFTWARE: Macintosh Wordperfect converted to PC ASCII Text
; SEQ ID NO 4
;   LENGTH: 2161
;   TYPE: DNA (cDNA)
;   ORGANISM: Nicotiana tabacum
US-09-001-826-4


  Query Match             70.5%;  Score 14.8;  DB 2;  Length 2161;
  Best Local Similarity   88.9%;  Pred. No. 70;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       4 tgaactccctgctacctt 21
           |||||| |||| ||||||
Db     593 tgaacttcctgatacctt 610


RESULT  13
US-09-001-826-25
; Sequence 25, Application US/09001826A
; Patent No. 5965727
; GENERAL INFORMATION:
;  APPLICANT: SONG, HEE-SOOK
;  APPLICANT: BROTHERTON, JEFFREY E.
;  APPLICANT: WIDHOLM, JACK M.
;  TITLE OF INVENTION: SELECTABLE MARKER AND PROMOTER FOR PLANT TISSUE CULTURE
;  TITLE OF INVENTION: TRANSFORMATION
;  FILE REFERENCE: UI001.C1
;  CURRENT APPLICATION NUMBER: US/09/001,826A
;  CURRENT FILING DATE: 1997-12-31
;  EARLIER APPLICATION NUMBER: 08/937,739;  60/025,140
;  EARLIER FILING DATE: 1997-07-25; 1996-07-26
;  NUMBER OF SEQ ID NOS: 31
;  SOFTWARE: Macintosh Wordperfect converted to PC ASCII Text
; SEQ ID NO 25
;   LENGTH: 2161
;   TYPE: DNA (cDNA)
;   ORGANISM: Nicotiana tabacum
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: LOCATION: 90..1940
US-09-001-826-25


  Query Match             70.5%;  Score 14.8;  DB 2;  Length 2161;
  Best Local Similarity   88.9%;  Pred. No. 70;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       4 tgaactccctgctacctt 21
           |||||| |||| ||||||
Db     593 tgaacttcctgatacctt 610


RESULT  14
US-08-366-547-1
; Sequence 1, Application US/08366547
; Patent No. 5843737
;  GENERAL INFORMATION:
;    APPLICANT:  Chen, Lan Bo
;    APPLICANT:  Bao, Shideng
;    TITLE OF INVENTION:  A NEW CANCER ASSOCIATED GENE, PROTEIN
;    TITLE OF INVENTION:  EXPRESSED THEREFROM AND USES THEREOF
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DAVID G. CONLIN; DIKE, BRONSTEIN, ROBERTS
;      ADDRESSEE:  & CUSHMAN
;      STREET:  130 Water Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  US
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/366,547
;      FILING DATE:  30-DEC-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Eisenstein, Ronald I.
;      REGISTRATION NUMBER:  30628
;      REFERENCE/DOCKET NUMBER:  45072
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 523-3400
;      TELEFAX:  (617) 523-6440
;      TELEX:  200291 STRE UR
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2652 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  120..2130
US-08-366-547-1
```

```
  Query Match           70.5%;  Score 14.8;  DB 2;  Length 2652;
  Best Local Similarity  88.9%;  Pred. No. 72;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 agtgaactccctgctacc 19
           |||||||| ||||||| |
Db    2037 AGTGAACTGCCTGCTAGC 2054


RESULT  15
US-08-276-151-1/c
; Sequence 1, Application US/08276151
; Patent No. 5597719
;  GENERAL INFORMATION:
;     APPLICANT:  Freed, Ellen
;     APPLICANT:  Ruggieri, Rosamaria
;     TITLE OF INVENTION:  Interaction of raf-1 and 14-3-3 Proteins
;     NUMBER OF SEQUENCES:  9
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Cooley Godward et al.
;       STREET:  Five Palo ALto Square
;       CITY:  Palo Alto
;       STATE:  CA
;       COUNTRY:  USA
;       ZIP:  94036
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/276,151
;       FILING DATE:  14-JUL-1994
;       CLASSIFICATION:  530
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Torchia, Ph.D., Timothy E
;       REGISTRATION NUMBER:  36,700
;       REFERENCE/DOCKET NUMBER:  ONYX-005/00US
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (415) 843-5481
;       TELEFAX:  (415) 857-0663
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2977 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA to mRNA
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
;     ORIGINAL SOURCE:
;       ORGANISM:  Homo sapiens
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  130..2076
US-08-276-151-1


  Query Match           70.5%;  Score 14.8;  DB 1;  Length 2977;
  Best Local Similarity  88.9%;  Pred. No. 74;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       4 tgaactccctgctacctt 21
           || ||| ||||||||||||
Db    2436 TGGACTGCCTGCTACCTT 2419


Search completed: July 25, 2001, 05:18:51
Job time: 9214 sec
```

```
Qy      3 tgctcctgccagaggtt 19
          |||||||||||||||||
Db    506 TGCTCCTGCCAGAGGTT 490


RESULT   5
AQ068059
LOCUS       AQ068059      356 bp    DNA             GSS       04-AUG-1998
DEFINITION  HS_2200_B2_F03_MF CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=2200 Col=6 Row=L, DNA sequence.
ACCESSION   AQ068059
VERSION     AQ068059.1  GI:3379327
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 356)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Sequence Tagged Connector
            Plate: 2200  row: L  column: 6
            Class: BAC ends
            High quality sequence stop: 356.
FEATURES             Location/Qualifiers
     source          1. .356
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=2200 Col=6 Row=L"
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
                     /sex="male"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E-Coli DH10B"
BASE COUNT      111 a     86 c     67 g     92 t
ORIGIN


  Query Match           86.3%;  Score 16.4;  DB 223;  Length 356;
  Best Local Similarity   94.4%;  Pred. No. 6.2e+02;
  Matches   17;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      2 ttgctcctgccagaggtt 19
          ||||| ||||||||||||
Db    106 TTGCTTCTGCCAGAGGTT 123


RESULT   6
AQ799987
LOCUS       AQ799987      530 bp    DNA             GSS       09-AUG-1999
DEFINITION  HS_5300_B2_G10_T7A RPCI-11 Human Male BAC Library Homo sapiens
            genomic clone Plate=876 Col=20 Row=N, DNA sequence.
ACCESSION   AQ799987
VERSION     AQ799987.1  GI:5717319
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 530)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
            or from Resear h Genetics (info@resgen.com). BAC end Web Server:
            http://www.htsc.washington.edu
            Plate: 876  row: N  column: 20
            Seq primer: T7
            Class: BAC ends
            High quality sequence stop: 530.
FEATURES             Location/Qualifiers
     source          1. .530
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=876 Col=20 Row=N"
                     /clone_lib="RPCI-11 Human Male BAC Library"
                     /sex="male"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     Male blood DNA was isolated from one randomly chosen donor
                     and partially digested with a combination of EcoRI and
                     EcoRI Methylase. Size selected DNA was cloned into the
                     pBACe3.6 vector at EcoRI sites"
BASE COUNT      109 a     120 c     167 g     130 t      4 others
ORIGIN


  Query Match           86.3%;  Score 16.4;  DB 233;  Length 530;
  Best Local Similarity   94.4%;  Pred. No. 6.6e+02;
  Matches   17;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 tttgctcctgccagaggt 18
          | ||||||||||||||||
Db    325 TGTGCTCCTGCCAGAGGT 342


RESULT   7
AI979827/c
LOCUS       AI979827      574 bp    mRNA            EST       10-JUL-2000
DEFINITION  pat.pk0004.f2 chicken activated T cell cDNA Gallus gallus cDNA
            clone pat.pk0004.f2 5' similar to FRUCTOSE-BISPHOSPHATE ALDOLASE B
            (LIVER-TYPE ALDOLASE), mRNA sequence.
ACCESSION   AI979827
VERSION     AI979827.1  GI:5882855
KEYWORDS    EST.
SOURCE      chicken.
  ORGANISM  Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.
REFERENCE   1  (bases 1 to 574)
  AUTHORS   Tirunagaru,V.G., Sofer,L., Cui,J. and Burnside,J.
  TITLE     An expressed sequence tag database of T-cell-enriched activated
            chicken splenocytes: sequence analysis of 5251 clones
  JOURNAL   Genomics 66 (2), 144-151 (2000)
  MEDLINE   20318616
COMMENT     Contact: Joan Burnside
            Molecular Endocrinology
            University of Delaware
            40 Townsend Hall, Newark, DE 19717, USA
            Tel: 302 831-1345
            Fax: 302 831-3651
```

Email: joan@UDel.Edu
Seq primer: T7.
FEATURES Location/Qualifiers
source 1. .574
/organism="Gallus gallus"
/db_xref="taxon:9031"
/clone="pat.pk0004.f2"
/clone_lib="chicken activated T cell cDNA"
/sex="male"
/cell_type="Con A-activated splenic T cell"
/lab_host="E.coli TOP10 F'"
/note="Vector: pcDNA3"
BASE COUNT 165 a 130 c 154 g 125 t
ORIGIN

Query Match 86.3%; Score 16.4; DB 104; Length 574;
Best Local Similarity 94.4%; Pred. No. 6.7e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 tttgctcctgccagaggt 18
|||| ||||||||||||||
Db 340 TTTGTTCCTGCCAGAGGT 323

RESULT 8
AZ664475
LOCUS AZ664475 616 bp DNA GSS 14-DEC-2000
DEFINITION 1M0544H08R Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC1M0544H08 R, DNA sequence.
ACCESSION AZ664475
VERSION AZ664475.1 GI:11801621
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 616)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D.,Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0544 row: H column: 08
Seq primer: CACACAGGAAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 616.
FEATURES Location/Qualifiers
source 1. .616
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0544H08"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptored DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to adaptored vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."
BASE COUNT 156 a 162 c 131 g 167 t
ORIGIN

Query Match 86.3%; Score 16.4; DB 247; Length 616;
Best Local Similarity 94.4%; Pred. No. 6.8e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 ttgctcctgccagaggtt 19
|||||||||| ||||||||
Db 319 TTGCTCCTGACAGAGGTT 336

RESULT 9
BE901069
LOCUS BE901069 680 bp mRNA EST 29-SEP-2000
DEFINITION 601674415F1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:3957315 5', mRNA sequence.
ACCESSION BE901069
VERSION BE901069.1 GI:10389877
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 680)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LLCM834 row: n column: 04
High quality sequence start: 2
High quality sequence stop: 608.
FEATURES Location/Qualifiers
source 1. .680
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3957315"
/clone_lib="NIH_MGC_21"
/tissue_type="choriocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: placenta; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT 117 a 190 c 214 g 159 t
ORIGIN

```
  Query Match           86.3%;  Score 16.4;  DB 141;  Length 680;
  Best Local Similarity  94.4%;  Pred. No. 6.9e+02;
  Matches   17;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 tttgctcctgccagaggt 18
          ||||||||||||||||| |
Db    371 TTTGCTCCTGCCAGAGCT 388


RESULT 10
BG109825/c
LOCUS       BG109825      978 bp    mRNA            EST       30-JAN-2001
DEFINITION  602280983F1 NIH_MGC_86 Homo sapiens cDNA clone IMAGE:4368405 5',
            mRNA sequence.
ACCESSION   BG109825
VERSION     BG109825.1  GI:12603331
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 978)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10022  row: f  column: 22
            High quality sequence stop: 751.
FEATURES             Location/Qualifiers
     source          1. .978
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4368405"
                     /clone_lib="NIH_MGC_86"
                     /tissue_type="osteosarcoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: bone; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 1.533 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT      250 a    250 c    276 g    202 t
ORIGIN


  Query Match           86.3%;  Score 16.4;  DB 173;  Length 978;
  Best Local Similarity  94.4%;  Pred. No. 7.3e+02;
  Matches   17;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 tttgctcctgccagaggt 18
          |||||||| |||||||||
Db    844 TTTGCTCCAGCCAGAGGT 827


RESULT 11
AI203923
LOCUS       AI203923      206 bp    mRNA            EST       28-OCT-1998
DEFINITION  qd72h07.x1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1735069
            3', mRNA sequence.
ACCESSION   AI203923
VERSION     AI203923.1  GI:3756529
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 206)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo
            , Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 1344   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 197.
FEATURES             Location/Qualifiers
     source          1. .206
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1735069"
                     /clone_lib="Soares_testis_NHT"
                     /sex="male"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was prepared from mRNA obtained from Clontech Laboratories
                     , Inc., and primed with  a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCCCAATTTTTTTTTTTTTTT 3']
                     Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization to Cot5, and was
                     constructed  by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT      50 a     48 c     53 g     55 t
ORIGIN


  Query Match           84.2%;  Score 16;  DB 17;  Length 206;
  Best Local Similarity  100.0%;  Pred. No. 9.1e+02;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      4 gctcctgccagaggtt 19
          ||||||||||||||||
Db     52 GCTCCTGCCAGAGGTT 67


RESULT 12
AA504455
LOCUS       AA504455      338 bp    mRNA            EST       18-AUG-1997
DEFINITION  aa60a05.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:825296 3',
            mRNA sequence.
ACCESSION   AA504455
VERSION     AA504455.1  GI:2240615
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 338)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
            Ph.D., Gerald Marti, M.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
```

DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 320.

FEATURES Location/Qualifiers
source 1. .338
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:825296"
/clone_lib="NCI_CGAP_GCB1"
/tissue_type="germinal center B cell"
/lab_host="DH10B"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was prepared from human tonsillar cells enriched for germinal center B cells by flow sorting (CD20+, IgD-), provided by Dr. Louis M. Staudt (NCI), Dr. David Allman (NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was primed with a Not I - oligo(dT) primer [5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCCTCATTTTTTTTTTTTTTTTT-3']. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 79 a 71 c 68 g 120 t
ORIGIN

Query Match 84.2%; Score 16; DB 8; Length 338;
Best Local Similarity 100.0%; Pred. No. 9.9e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 gctcctgccagaggtt 19
||||||||||||||||
Db 86 GCTCCTGCCAGAGGTT 101

RESULT 13
AA577198
LOCUS AA577198 341 bp mRNA EST 12-SEP-1997
DEFINITION nm87b11.s1 NCI_CGAP_Co9 Homo sapiens cDNA clone IMAGE:1075197 3', mRNA sequence.
ACCESSION AA577198
VERSION AA577198.1 GI:2354672
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 341)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Ilan Kirsch, M.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 919 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 181.

FEATURES Location/Qualifiers
source 1. .341
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1075197"
/clone_lib="NCI_CGAP_Co9"
/tissue_type="colon tumor RER+"
/lab_host="DH10B"
/note="Organ: colon; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from RER+ colon tumor, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library is not normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo (Soares4)."

BASE COUNT 79 a 85 c 82 g 95 t
ORIGIN

Query Match 84.2%; Score 16; DB 9; Length 341;
Best Local Similarity 100.0%; Pred. No. 9.9e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 gctcctgccagaggtt 19
||||||||||||||||
Db 189 GCTCCTGCCAGAGGTT 204

RESULT 14
AQ751483
LOCUS AQ751483 738 bp DNA GSS 19-JUL-1999
DEFINITION HS_5576_B2_B02_SP6 RPCI-11 Human Male BAC Library Homo sapiens genomic clone Plate=1152 Col=4 Row=D, DNA sequence.
ACCESSION AQ751483
VERSION AQ751483.1 GI:5538641
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 738)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm) or from Resear h Genetics (info@resgen.com). BAC end Web Server: http://www.htsc.washington.edu
Plate: 1152 row: D column: 4
Seq primer: SP6
Class: BAC ends
High quality sequence stop: 738.

FEATURES Location/Qualifiers
source 1. .738
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=1152 Col=4 Row=D"
/clone_lib="RPCI-11 Human Male BAC Library"
/sex="male"
/note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI; Male blood DNA was isolated from one randomly chosen donor

and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACe3.6 vector at EcoRI sites"
BASE COUNT 178 a 167 c 190 g 203 t
ORIGIN

Query Match 84.2%; Score 16; DB 233; Length 738;
Best Local Similarity 100.0%; Pred. No. 1.1e+03;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 gctcctgccagaggtt 19
||||||||||||||||
Db 197 GCTCCTGCCAGAGGTT 212

RESULT 15
W25430/c
LOCUS W25430 115 bp mRNA EST 10-OCT-1996
DEFINITION zb90c01.r1 Soares_senescent_fibroblasts_NbHSF Homo sapiens cDNA clone IMAGE:310848 5', mRNA sequence.
ACCESSION W25430
VERSION W25430.1 GI:1303285
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 115)
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and Wilson,R.
TITLE The WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 627 Std Error: 0.00
Seq primer: mob.REGA+ET.
FEATURES Location/Qualifiers
source 1. .115
/organism="Homo sapiens"
/db_xref="GDB:1254262"
/db_xref="taxon:9606"
/clone="IMAGE:310848"
/clone_lib="Soares_senescent_fibroblasts_NbHSF"
/tissue_type="senescent fibroblast"
/lab_host="DH10B (ampicillin resistant)"
/note="Vector: pT7T3D (Pharmacia) with a modified polylinker V_TYPE: phagemid; Site_1: Not I; Site_2: Eco RI ; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGGAGCGGCCGCATTTTTTTTTTTTTTTTTT 3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT7T3 vector (Pharmacia). Library went through one round of normalization to a Cot - 5. Library constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT 40 a 17 c 27 g 31 t
ORIGIN

Query Match 83.2%; Score 15.8; DB 190; Length 115;
Best Local Similarity 89.5%; Pred. No. 1.1e+03;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 tttgctcctgccagaggtt 19
||| |||||||||||| ||
Db 49 TTTACTCCTGCCAGAGCTT 31

Search completed: July 25, 2001, 04:54:56
Job time: 10429 sec

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: July 25, 2001, 05:23:04 ; Search time 247.87 Seconds
(without alignments)
50.664 Million cell updates/sec

Title: US-09-142-095-3
Perfect score: 20
Sequence: 1 gtcacgtgacacagtcaaac 20

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_0601:*
1: /SIDS1/gcgdata/geneseq/geneseqn/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	20	100.0	20	18	AAT79543	UGT1*1 gene exon 1
2	20	100.0	620	18	AAT79540	Upstream DNA seque
c 3	16.4	82.0	22	19	AAV13057	3'-flanking region
4	15.8	79.0	27	21	AAZ57627	Collagen IV alpha
5	15.8	79.0	27	21	AAZ48039	Human Alport syndr
c 6	15.8	79.0	300	20	AAX98515	Human cancer cell
7	15.8	79.0	392	21	AAA50369	Human Goodpasture
8	15.8	79.0	507	21	AAA50370	Human Goodpasture
9	15.8	79.0	680	21	AAA50368	Human Goodpasture
10	15.8	79.0	685	21	AAA50367	Human Goodpasture
11	15.8	79.0	738	21	AAZ57158	Human type IV coll
12	15.8	79.0	900	20	AAZ20091	Type IV collagen N
13	15.8	79.0	900	21	AAA90993	Human alpha3(IV)NC
c 14	15.2	76.0	1276	21	AAF13921	Aspergillus oryzae
15	15.2	76.0	1801	21	AAC35997	Arabidopsis thalia
16	15.2	76.0	1803	21	AAC51372	Arabidopsis thalia
c 17	15	75.0	2530	21	AAC39424	Arabidopsis thalia
c 18	14.8	74.0	393	22	AAF66883	Novel human polynu
19	14.8	74.0	430	16	AAT19914	Human gene signatu
c 20	14.8	74.0	519	19	AAV69028	DNA molecule encod
c 21	14.8	74.0	519	21	AAC80982	Human breast tumou
c 22	14.8	74.0	887	7	AAN60894	Sequence of plasmi
c 23	14.8	74.0	1100	5	AAN40127	Sequence encoding
c 24	14.8	74.0	1100	5	AAN40104	Sequence of human
c 25	14.8	74.0	1101	4	AAN30003	Sequence including
c 26	14.8	74.0	1101	5	AAN40133	Sequence of human
c 27	14.8	74.0	1101	5	AAN40240	Human immunointerf
c 28	14.8	74.0	1144	4	AAN30109	Sequence encoding
c 29	14.8	74.0	1156	6	AAN50276	Human recombinant
c 30	14.8	74.0	1161	6	AAN50327	Human interferon-g
c 31	14.8	74.0	1161	7	AAN60103	Interferon gamma i
c 32	14.8	74.0	1210	4	AAN30055	Sequence of the pl
c 33	14.8	74.0	1226	13	AAQ24919	Human interferon-g
c 34	14.8	74.0	1350	6	AAN50335	Sequence of the IF
35	14.8	74.0	2059	7	AAN60895	Sequence of plasmi
36	14.8	74.0	2171	22	AAF54722	Nucleotide sequenc
37	14.8	74.0	2724	22	AAF54721	Nucleotide sequenc
38	14.8	74.0	2740	16	AAQ85355	Prosaposin cDNA.
39	14.8	74.0	2749	20	AAX08488	Human prosaposin c
c 40	14.8	74.0	2827	7	AAN60893	Sequence of plasmi
41	14.8	74.0	3000	7	AAN60892	Sequence of plasmi
c 42	14.8	74.0	5274	21	AAC42949	Arabidopsis thalia
c 43	14.8	74.0	5319	21	AAC50069	Arabidopsis thalia
c 44	14.8	74.0	5961	16	AAQ14060	Human interferon-g
45	14.8	74.0	68940	20	AAX57351	Human chromosome 6

ALIGNMENTS

RESULT 1
AAT79543
ID AAT79543 standard; DNA; 20 BP.
XX
AC AAT79543;
XX
DT 23-JAN-1998 (first entry)
XX
DE UGT1*1 gene exon 1 upstream PCR primer C.
XX
KW Uridine diphosphate glucuronosyltransferase gene; UGT;
KW Gilbert's syndrome; GS; unconjugated hyperbilirubinaemia;
KW bilirubin glucuronidation; Crigler-Najjar; type 2; drug metabolism;
KW Drug trial efficiency; screening; PCR primer; ss.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN WO9732042-A2.
XX
PD 04-SEP-1997.
XX
PF 03-MAR-1997; 97WO-GB00577.
XX
PR 16-MAR-1996; 96GB-0005598.
PR 01-MAR-1996; 96GB-0004480.
XX
PA (UYDU-) UNIV DUNDEE.
XX
PI Burchell B;
XX
DR WPI; 1997-448702/41.
XX
PT Improving drug trial efficiency comprises identifying participants

```
PT   with Gilbert's syndrome - useful as their altered drug metabolism
PT   may hinder result interpretation
XX
PS   Claim 14; Page 12; 31pp; English.
XX
CC   This PCR primer (with primers AAT79542-44) flanks the TATA box sequence
CC   upstream of the uridine diphosphate glucuronosyltransferase (UGT) gene
CC   1*1 exon 1 (see AAT79540), and was used to amplify fragments of 98-100
CC   bp. This gene is known to be associated with Gilbert's syndrome (GS). GS
CC   is a mild, common form of unconjugated hyperbilirubinaemia associated
CC   with reduced bilirubin glucuronidation capacity. Analysis of the genetic
CC   basis of GS has allowed 2 forms to be identified. One is a mild form
CC   associated with a homozygous 2 bp insertion in the TATA sequence
CC   upstream of the UGT1*1 exon 1, and the other is a more severe form
CC   associated with heterozygosity for a mutation which, when homozygous,
CC   causes Crigler-Najjar type 2 disease. The first form is autosomal
CC   recessive and the second is inherited dominantly. Patients suffering from
CC   GS, which is benign, may have altered metabolism of some drugs, making it
CC   difficult to determine if an effect is due to the drug or the syndrome.
CC   Drug trial efficiency would be improved if potential participants can be
CC   screened for the genetic basis of GS, and eliminated or included  on
CC   basis of them possessing or not possessing GS.
XX
SQ   Sequence 20 BP; 7 A; 6 C; 4 G; 3 T; 0 other;


  Query Match             100.0%;  Score 20;  DB 18;  Length 20;
  Best Local Similarity   100.0%;  Pred. No. 0.11;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gtcacgtgacacagtcaaac 20
          ||||||||||||||||||||
Db      1 gtcacgtgacacagtcaaac 20


RESULT   2
AAT79540
ID   AAT79540 standard; DNA; 620 BP.
XX
AC   AAT79540;
XX
DT   23-JAN-1998  (first entry)
XX
DE   Upstream DNA sequence of UGT1*1 gene exon 1.
XX
KW   Uridine diphosphate glucuronosyltransferase gene; UGT;
KW   Gilbert's syndrome; GS; unconjugated hyperbilirubinaemia;
KW   bilirubin glucuronidation; Crigler-Najjar; type 2; drug metabolism;
KW   Drug trial efficiency; screening; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   protein_bind     2..8
FT                    /*tag=  a
FT                    /bound_moeity= AP1
FT   GC_signal        93..102
FT                    /*tag=  b
FT                    /note= "SP1 binding site"
FT   misc_feature     180..191
FT                    /*tag=  c
FT                    /note= "feature indicated in patent, but no further
FT                          explanation is given"
FT   protein_bind     316..323
FT                    /*tag=  d
FT                    /bound_moeity= AP3
FT   protein_bind     317..324
FT                    /*tag=  e
FT                    /bound_moeity= CLBP
FT   misc_feature     366..376
FT                    /*tag=  f
FT                    /note= "feature indicated in patent, but no further
FT                          explanation is given"
FT   protein_bind     386..392
FT                    /*tag=  g
FT                    /bound_moeity= AP1
FT   protein_bind     513..519
FT                    /*tag=  h
FT                    /bound_moeity= AP1
FT   misc_feature     513..520
FT                    /*tag=  i
FT                    /note= "feature indicated in patent, but no further
FT                          explanation is given"
FT   protein_bind     520..532
FT                    /*tag=  j
FT                    /bound_moeity= HNF1
FT   TATA_signal      558..572
FT                    /*tag=  k
FT                    /note= "corresponds to positions -53 to -39 in patent"
FT   misc_feature     595
FT                    /*tag=  l
FT                    /note= "feature indicated in patent, but no further
FT                          explanation is given"
FT   CDS              612
FT                    /*tag=  m
FT                    /codon_start= 612
XX
PN   WO9732042-A2.
XX
PD   04-SEP-1997.
XX
PF   03-MAR-1997;   97WO-GB00577.
XX
PR   16-MAR-1996;   96GB-0005598.
PR   01-MAR-1996;   96GB-0004480.
XX
PA   (UYDU-) UNIV DUNDEE.
XX
PI   Burchell B;
XX
DR   WPI; 1997-448702/41.
XX
PT   Improving drug trial efficiency comprises identifying participants
PT   with Gilbert's syndrome - useful as their altered drug metabolism
PT   may hinder result interpretation
XX
PS   Claim 11; Fig 4; 31pp; English.
XX
CC   This sequence represents the upstream sequence, positions -611 to 9
CC   in the patent, of uridine diphosphate glucuronosyltransferase (UGT)
CC   gene 1*1 exon 1. This gene is known to be associated with Gilbert's
CC   syndrome (GS). GS is a mild, common form of unconjugated
CC   hyperbilirubinaemia associated with reduced bilirubin glucuronidation
CC   capacity. Analysis of the genetic basis of GS has allowed 2 forms to be
CC   identified. One is a mild form associated with a homozygous 2 bp
CC   insertion in the TATA sequence upstream of the UGT1*1 exon 1, and
CC   the other is a more severe form associated with heterozygosity for a
CC   mutation which, when homozygous, causes Crigler-Najjar type 2 disease.
CC   The first form is autosomal recessive and the second is inherited
CC   dominantly. Patients suffering from GS, which is benign, may have
CC   altered metabolism of some drugs, making it difficult to determine if an
CC   effect is due to the drug or the syndrome. Drug trial efficiency would
CC   be improved if potential participants can be screened for the genetic
CC   basis of GS, and eliminated or included  on basis of them possessing or
CC   not possessing GS. In this case, screening involves PCR amplification
CC   of the UGT gene, using the primers described in AAT79541-44.
XX
SQ   Sequence 620 BP; 157 A; 127 C; 151 G; 185 T; 0 other;


  Query Match             100.0%;  Score 20;  DB 18;  Length 620;
  Best Local Similarity   100.0%;  Pred. No. 0.19;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gtcacgtgacacagtcaaac 20
          ||||||||||||||||||||
```

```
Db    508 gtcacgtgacacagtcaaac 527


RESULT   3
AAV13057/c
ID   AAV13057 standard; DNA; 22 BP.
XX
AC   AAV13057;
XX
DT   18-MAY-1998  (first entry)
XX
DE   3'-flanking region of an interleukin 6 gene PCR primer.
XX
KW   Interleukin 6; IL-6; PCR primer; genotype; osteoporosis; human;
KW   polymorphism; ss.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   WO9743446-A2.
XX
PD   20-NOV-1997.
XX
PF   16-MAY-1997;   97WO-GB01337.
XX
PR   16-MAY-1996;   96GB-0010281.
XX
PA   (GEMI-) GEMINI INT HOLDINGS LTD.
XX
PI   Grant SFA,  Ralston SH;
XX
DR   WPI; 1998-008905/01.
XX
PT   Determination of genotype of an IL-6 gene in an individual - used
PT   for diagnosis of or identification of a predisposition or
PT   susceptibility to a disease, particularly osteoporosis
XX
PS   Claim 9; Page 6; 20pp; English.
XX
CC   The present sequence represents a PCR primer used in the present
CC   invention. The present invention describes an in vitro method of
CC   diagnosis of a disease in an animal. The method comprises determining
CC   the genotype of an interleukin-6 (IL-6) gene in the animal. The
CC   invention also describes: (1) a method of identifying an animal
CC   predisposed or susceptible to a disease, comprising determining the
CC   genotype of an IL-6 gene in the animal; (2) a composition for use in
CC   diagnosing a disease in an animal, the composition comprising one or
CC   more primer nucleic acid molecules adapted to amplify a portion of a
CC   3' flanking region of an IL-6 gene in the animal; and (3) a composition
CC   for use in identifying an animal predisposed or susceptible to a
CC   disease, the composition comprising one or more primer nucleic acid
CC   molecules adapted to amplify a portion of a 3' flanking region of an
CC   IL-6 gene in the animal. The method can be used for the diagnosis of
CC   or identifying predisposition or susceptibility to a disease,
CC   particularly osteoporosis.
XX
SQ   Sequence 22 BP; 4 A; 4 C; 7 G; 7 T; 0 other;


  Query Match            82.0%;  Score 16.4;  DB 19;  Length 22;
  Best Local Similarity  94.4%;  Pred. No. 9;
  Matches   17;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      2 tcacgtgacacagtcaaa 19
          ||||||||||||| |||||
Db     19 TCACGTGACACACTCAAA 2


RESULT   4
AAZ57627
ID   AAZ57627 standard; DNA; 27 BP.
XX
AC   AAZ57627;
XX
DT   05-APR-2000  (first entry)
XX
DE   Collagen IV alpha 3 deletion mutant fragment #1.
XX
KW   Collagen IV alpha 3; deletion mutant; nucleotide probe; genomic profile;
KW   Alport syndrome; disease prognosis; health screening;
KW   health management tool; ss.
XX
OS   Unidentified.
XX
PN   WO9964626-A2.
XX
PD   16-DEC-1999.
XX
PF   04-JUN-1999;   99WO-GB01779.
XX
PR   06-JUN-1998;   98GB-0012098.
PR   23-DEC-1998;   98GB-0028289.
XX
PA   (GENO-) GENOSTIC PHARMA LTD.
XX
PI   Roberts GW;
XX
DR   WPI; 2000-097546/08.
XX
PT   Nucleotide probes used in genetic screens for determining genomic
PT   profiles, e.g. for prognosis or management -
XX
PS   Example 5; Page 25; 149pp; English.
XX
CC   This sequence is an collagen IV alpha 3 deletion mutant fragment. The
CC   mutated gene is an example of a variant that can be detected using the
CC   new nucleotide probes of the invention. The probes consist of
CC   complementary DNA or RNA for the detection of variants in a group of
CC   target genes associated with particular diseases. The mutant variant
CC   represented by this sequence gives rise to Alport syndrome phenotype. The
CC   specification lists approximately 2500 genes that are target groups for
CC   the nucleotide probes of the invention. The probes are used to determine
CC   the genomic profile of an individual, particularly for;
CC   (a) prognosis and management of disease (or determining susceptibility
CC   to disease);
CC   (b) predicting a patient's response to therapy and symptom profiles;
CC   (c) for health screening;
CC   (d) to develop new strategies for therapy and clinical trials; and
CC   (e) to construct health management algorithms or models, e.g. of the
CC   likelihood of developing disease.
CC   Antibodies specific for the proteins encoded by the genes in the target
CC   list, can be used similarly. Identification of a core group of genes
CC   associated with disease, makes possible the use of genetic profiling as a
CC   universal health management tool.
XX
SQ   Sequence 27 BP; 7 A; 10 C; 3 G; 7 T; 0 other;


  Query Match            79.0%;  Score 15.8;  DB 21;  Length 27;
  Best Local Similarity  89.5%;  Pred. No. 19;
  Matches   17;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      2 tcacgtgacacagtcaaac 20
          ||||  |||||||||||||
Db      8 tcacccgacacagtcaaac 26


RESULT   5
AAZ48039
ID   AAZ48039 standard; DNA; 27 BP.
XX
AC   AAZ48039;
XX
DT   08-MAR-2000  (first entry)
XX
DE   Human Alport syndrome oligonucleotide #1.
```

XX
KW Human; glycogen storage disease 2; genetic filing; identification;
KW polymorphic variation; mutation; detection; probe; prognosis; headache;
KW gene associated disease; health screening; therapy; clinical trial;
KW absorption; distribution; metabolism; elimination; oncology; dementia;
KW central nervous system; behaviour; brain injury; psychosis; personality;
KW cardiovascular; gastrointestinal; respiratory; immunology; development;
KW skin; bone; muscle; endocrine; sexual dysfunction; ss.
XX
OS Homo sapiens.
XX
PN WO9964627-A2.
XX
PD 16-DEC-1999.
XX
PF 04-JUN-1999; 99WO-GB01780.
XX
PR 06-JUN-1998; 98GB-0012099.
PR 20-JUN-1998; 98GB-0013291.
PR 24-JUN-1998; 98GB-0013611.
PR 27-JUN-1998; 98GB-0013835.
PR 01-JUL-1998; 98GB-0014110.
PR 07-JUL-1998; 98GB-0014580.
PR 16-JUL-1998; 98GB-0015438.
PR 18-JUL-1998; 98GB-0015574.
PR 18-JUL-1998; 98GB-0015576.
PR 24-JUL-1998; 98GB-0016085.
PR 24-JUL-1998; 98GB-0016086.
PR 05-AUG-1998; 98GB-0016921.
PR 07-AUG-1998; 98GB-0017097.
PR 08-AUG-1998; 98GB-0017200.
PR 14-AUG-1998; 98GB-0017632.
PR 19-AUG-1998; 98GB-0017943.
XX
PA (GENO-) GENOSTIC PHARMA LTD.
XX
PI Roberts GW;
XX
DR WPI; 2000-097547/08.
XX
PT Nucleotide probes used in genetic screens for determining genomic
PT profiles, e.g. for prognosis or management -
XX
PS Example 5; Page 25; 745pp; English.
XX
CC The present invention describes new nucleotide probes (I) comprising
CC complementary DNA or RNA, used in the detection of variants in a group
CC of target genes associated with particular diseases. Detected variants
CC are mutations or polymorphisms. Included in the specification are lists
CC of genes for the following target groups: (i) ADME (absorption,
CC distribution, metabolism and elimination - about 1000 genes);
CC (ii) oncology (about 1200 genes); (iii) central nervous system (about
CC 1100 genes); (iv) behavioural (about 250 genes); (v) brain injury (about
CC 850 genes); (vi) dementia (about 700 genes); (vii) psychoses and
CC personality (about 500 genes); (viii) cardiovascular (about 1100 genes);
CC (ix) gastrointestinal (about 1100 genes); (x) respiratory (about 800
CC genes); (xi) immunology (about 1100 genes); (xii) development (about 2000
CC and endocrine (about 1100 genes); (xiv) metabolism
CC (xvi) sexual dysfunction (about 250 genes); (xv) headache (about 250 genes); and
CC genomic profile of an individual, particularly for; (a) prognosis and
CC management of disease (or determining susceptibility to disease);
CC (b) predicting a patient's response to therapy and symptom profiles;
CC (c) for health screening; (d) to develop new strategies for therapy and
CC clinical trials; and (e) to construct health management algorithms or
CC models, e.g. of the likelihood of developing disease. The present
CC sequence represents an oligonucleotide used to illustrate a deletion in
CC Alport syndrome, which is given in an example from the present invention.
XX
SQ Sequence 27 BP; 7 A; 10 C; 3 G; 7 T; 0 other;

Query Match 79.0%; Score 15.8; DB 21; Length 27;

Best Local Similarity 89.5%; Pred. No. 19;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 tcacgtgacacagtcaaac 20
|||| ||||||||||||
Db 8 tcacccgacacagtcaaac 26

RESULT 6
AAX98515/c
ID AAX98515 standard; cDNA; 300 BP.
XX
AC AAX98515;
XX
DT 24-SEP-1999 (first entry)
XX
DE Human cancer cell derived cDNA #241.
XX
KW Cancer; human; colon; breast; lung; transmembrane receptor; ATPase;
KW integral membrane protein; aspartyl protease; GATA family; wnt family;
KW transcription factor; G-protein alpha subunit; protein phosphatase;
KW phorbolester binding protein; diacylglycerol binding protein; trypsin;
KW protein kinase; tyrosine phosphatase; developmental signalling protein;
KW WW/rsp5/WWP domain; therapy; forensic; genetic mapping; diagnostic;
KW detection; treatment; cervical; melanoma; colorectal adenocarcinoma;
KW Wilm's tumour; retinoblastoma; sarcoma; myosarcoma; lung carcinoma;
KW leukemia; lymphoma; dysplasia; hyperplasia; endometrium; adrenal;
KW prostate; ss.
XX
OS Homo sapiens.
XX
PN WO9933982-A2.
XX
PD 08-JUL-1999.
XX
PF 22-DEC-1998; 98WO-US27610.
XX
PR 21-DEC-1998; 98US-0217471.
PR 23-DEC-1997; 97US-0068755.
PR 03-APR-1998; 98US-0080664.
PR 21-OCT-1998; 98US-0105234.
PR 27-OCT-1998; 98US-0105877.
XX
PA (CHIR) CHIRON CORP.
PA (HYSE-) HYSEQ INC.
XX
PI Crkvenjakov R, Dickson M, Drmanac R, Drmanac S;
PI Escobedo J, Garcia PD, Garcia V, Giese K, Innis MA;
PI Jones LW, Kassam A, Kennedy GC, Kita D, Labat I;
PI Lamson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;
PI Stache-Crain B, Sudduth-Klinger J, Williams LT;
XX
DR WPI; 1999-430243/36.
XX
PT New isolated human polynucleotides
XX
PS Claim 1; Page 375; 591pp; English.
XX
CC This invention describes novel isolated human polynucleotides obtained
CC by screening for differential expression in colon cancer, breast cancer
CC and lung cancer cell lines. The polynucleotides of the invention are
CC represented in AAX98275-X99118 and encode polypeptides of protein
CC families selected from 4 transmembrane segments integral membrane
CC proteins, 7 transmembrane receptors, ATPases associated with various
CC cellular activities (AAA), eukaryotic aspartyl proteases, GATA family of
CC transcription factors, G-protein alpha subunit, phorbolesters or
CC diacylglycerol binding proteins, protein kinase, protein phosphatase 2C,
CC protein tyrosine phosphatase, trypsin, wnt family of developmental
CC signalling proteins and WW/rsp5/WWP domain containing proteins. The
CC encoded polypeptides also have a functional domain selected from Ank
CC repeat, basic region plus leucine zipper transcription factors,
CC bromodomain, EF-hand, SH3 domain, WD domain/G-beta repeats, zinc finger
CC (C2H2 type), zinc finger (CCHC class), and zinc-binding metalloprotease

CC domain. The polynucleotides encode polypeptides with similarity to known
CC protein families and are predicted to have similar properties. The novel
CC polynucleotides can be used to develop products for use as therapeutic
CC agents and in forensics, genetic analysis, mapping and diagnostic
CC applications. In particular, the product can be used for the detection
CC and management of cancers. They can be used for treating e.g, cervical
CC cancers, melanomas, colorectal adenocarcinomas, Wilm's tumour, sarcomas,
CC retinoblastoma, myosarcomas, lung carcinomas, leukemias, such as chronic
CC myelogenous leukemia, promyelocytic leukemia, monocytic leukemia, and
CC myeloid leukemia, and lymphomas such as histiocytic lymphoma, anhydric
CC hereditary ectodermal dysplasia, congenital alveolar dysplasia,
CC epithelial dysplasia of the cervix, fibrous dysplasia of bone, and
CC mammary dysplasia, hyperplasias, e.g. endometrial, adrenal, breast,
CC prostate or thyroid hyperplasias or pseudoepitheliomatous hyperplasia of
CC the skin.
XX
SQ Sequence 300 BP; 65 A; 110 C; 40 G; 85 T; 0 other;

Query Match 79.0%; Score 15.8; DB 20; Length 300;
Best Local Similarity 89.5%; Pred. No. 28;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtcacgtgacacagtcaaa 19
|||| |||||| |||||||
Db 206 GTCAGGTGACAGAGTCAAA 188

RESULT 7
AAA50369
ID AAA50369 standard; DNA; 392 BP.
XX
AC AAA50369;
XX
DT 20-NOV-2000 (first entry)
XX
DE Human Goodpasture antigen DeltaIII/IV/V cDNA.
XX
KW Goodpasture antigen; GPDeltaIII/IV/V; human; GPBP;
KW goodpasture antigen binding protein; autoimmune disease; apoptosis;
KW cancer; tumour; gene therapy; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1..210
FT /*tag= a
FT /product= Met-deleted GPDeltaIII/IV/V
FT /partial
XX
PN WO200050607-A2.
XX
PD 31-AUG-2000.
XX
PF 24-FEB-2000; 2000WO-IB00324.
XX
PR 24-FEB-1999; 99US-0121483.
XX
PA (SAUS/) SAUS J.
XX
PI Saus J;
XX
DR WPI; 2000-572094/53.
DR P-PSDB; AAY95920.
XX
PT Novel Goodpasture antigen binding proteins useful for diagnosing and
PT treating autoimmune disorders, tumor, and preventing cell apoptosis -
XX
PS Claim 25; Page 153-154; 158pp; English.
XX
CC The present sequence is that of cDNA encoding human Goodpasture
CC antigen (GP) DeltaIII/IV/V (see AAY95920), an alternative form of
CC human GP resulting from splicing out of exons III, IV and V. The
CC cDNA was obtained by subcloning a cDNA encoding the protein into a
CC modified pET15b vector including an initiator Met. The invention
CC relates to novel Goodpasture antigen binding proteins (GPBPs, see
CC AAY95900-11), which bind to and phosphorylate the unique N-terminal
CC region of human GP, and which are highly expressed in several
CC autoimmune conditions. Claimed methods for treating an autoimmune
CC disorder, cell apoptosis or a tumour involve modifying the
CC expression or activity of GPBP, especially using a GP-derived
CC peptide, such as GPDeltaIII/IV/V or a nucleic acid sequence
CC encoding it.
XX
SQ Sequence 392 BP; 134 A; 80 C; 81 G; 97 T; 0 other;

Query Match 79.0%; Score 15.8; DB 21; Length 392;
Best Local Similarity 89.5%; Pred. No. 29;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 tcacgtgacacagtcaaac 20
|||| ||||||||||||
Db 65 tcacccgacacagtcaaac 83

RESULT 8
AAA50370
ID AAA50370 standard; DNA; 507 BP.
XX
AC AAA50370;
XX
DT 20-NOV-2000 (first entry)
XX
DE Human Goodpasture antigen DeltaIII/V cDNA.
XX
KW Goodpasture antigen; GPDeltaIII/V; human; GPBP;
KW goodpasture antigen binding protein; autoimmune disease; apoptosis;
KW cancer; tumour; gene therapy; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1..219
FT /*tag= a
FT /product= Met-deleted GPDeltaIII/V
FT /partial
XX
PN WO200050607-A2.
XX
PD 31-AUG-2000.
XX
PF 24-FEB-2000; 2000WO-IB00324.
XX
PR 24-FEB-1999; 99US-0121483.
XX
PA (SAUS/) SAUS J.
XX
PI Saus J;
XX
DR WPI; 2000-572094/53.
DR P-PSDB; AAY95921.
XX
PT Novel Goodpasture antigen binding proteins useful for diagnosing and
PT treating autoimmune disorders, tumor, and preventing cell apoptosis -
XX
PS Claim 25; Page 154-155; 158pp; English.
XX
CC The present sequence is that of cDNA encoding human Goodpasture
CC antigen (GP) DeltaIII/V (see AAY95921), an alternative form of
CC human GP resulting from splicing out of exons III and V. The
CC cDNA was obtained by subcloning a cDNA encoding the protein into a
CC modified pET15b vector including an initiator Met. The invention
CC relates to novel Goodpasture antigen binding proteins (GPBPs, see
CC AAY95900-11), which bind to and phosphorylate the unique N-terminal
CC region of human GP, and which are highly expressed in several

CC autoimmune conditions. Claimed methods for treating an autoimmune
CC disorder, cell apoptosis or a tumour involve modifying the
CC expression or activity of GPBP, especially using a GP-derived
CC peptide, such as GPDeltaIII/V or a nucleic acid sequence encoding
CC it.
XX
SQ Sequence 507 BP; 159 A; 113 C; 104 G; 131 T; 0 other;

Query Match 79.0%; Score 15.8; DB 21; Length 507;
Best Local Similarity 89.5%; Pred. No. 31;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 tcacgtgacacagtcaaac 20
|||| ||||||||||||
Db 65 tcacccgacacagtcaaac 83

RESULT 9
AAA50368
ID AAA50368 standard; DNA; 680 BP.
XX
AC AAA50368;
XX
DT 20-NOV-2000 (first entry)
XX
DE Human Goodpasture antigen DeltaIII cDNA.
XX
KW Goodpasture antigen; GPDeltaIII; human; GPBP;
KW goodpasture antigen binding protein; autoimmune disease; apoptosis;
KW cancer; tumour; gene therapy; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1..219
FT /*tag= a
FT /product= Met-deleted GPDeltaIII
FT /partial
XX
PN WO200050607-A2.
XX
PD 31-AUG-2000.
XX
PF 24-FEB-2000; 2000WO-IB00324.
XX
PR 24-FEB-1999; 99US-0121483.
XX
PA (SAUS/) SAUS J.
XX
PI Saus J;
XX
DR WPI; 2000-572094/53.
DR P-PSDB; AAY95919.
XX
PT Novel Goodpasture antigen binding proteins useful for diagnosing and
PT treating autoimmune disorders, tumor, and preventing cell apoptosis
XX
PS Claim 25; Page 152-153; 158pp; English.
XX
CC The present sequence is that of cDNA encoding human Goodpasture
CC antigen (GP) DeltaIII (see AAY95919), i.e. an alternative form of
CC human GP resulting from splicing out of exon III. The cDNA was
CC obtained by subcloning a cDNA encoding the protein into a modified
CC pET15b vector including an initiator Met. The invention relates
CC to novel Goodpasture antigen binding proteins (GPBPs, see
CC AAY95900-11), which bind to and phosphorylate the unique N-terminal
CC region of human GP, and which are highly expressed in several
CC autoimmune conditions. Claimed methods for treating an autoimmune
CC disorder, cell apoptosis or a tumour involve modifying the
CC expression or activity of GPBP, especially using a GP-derived
CC peptide, such as GPDeltaIII or a nucleic acid sequence encoding it.
XX
SQ Sequence 680 BP; 204 A; 159 C; 145 G; 172 T; 0 other;

Query Match 79.0%; Score 15.8; DB 21; Length 680;
Best Local Similarity 89.5%; Pred. No. 32;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 tcacgtgacacagtcaaac 20
|||| ||||||||||||
Db 65 tcacccgacacagtcaaac 83

RESULT 10
AAA50367
ID AAA50367 standard; DNA; 685 BP.
XX
AC AAA50367;
XX
DT 20-NOV-2000 (first entry)
XX
DE Human Goodpasture antigen DeltaV cDNA.
XX
KW Goodpasture antigen; GPDeltaV; goodpasture antigen binding protein;
KW GPBP; human; autoimmune disease; apoptosis; cancer; tumour;
KW gene therapy; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1..636
FT /*tag= a
FT /product= Met-deleted GPDeltaV
FT /partial
XX
PN WO200050607-A2.
XX
PD 31-AUG-2000.
XX
PF 24-FEB-2000; 2000WO-IB00324.
XX
PR 24-FEB-1999; 99US-0121483.
XX
PA (SAUS/) SAUS J.
XX
PI Saus J;
XX
DR WPI; 2000-572094/53.
DR P-PSDB; AAY95918.
XX
PT Novel Goodpasture antigen binding proteins useful for diagnosing and
PT treating autoimmune disorders, tumor, and preventing cell apoptosis
XX
PS Claim 25; Page 150-151; 158pp; English.
XX
CC The present sequence is that of cDNA encoding human Goodpasture
CC antigen (GP) DeltaV (see AAY95918), i.e. an alternative form of
CC human GP resulting from splicing out of exon V. The cDNA was
CC obtained by subcloning a cDNA encoding the protein into a modified
CC pET15b vector including an initiator Met. The invention relates
CC to novel Goodpasture antigen binding proteins (GPBPs, see
CC AAY95900-11), which bind to and phosphorylate the unique N-terminal
CC region of human GP, and which are highly expressed in several
CC autoimmune conditions. Claimed methods for treating an autoimmune
CC disorder, cell apoptosis or a tumour involve modifying the
CC expression or activity of GPBP, especially using a GP-derived
CC peptide, such as GPDeltaV or a nucleic acid sequence encoding it.
XX
SQ Sequence 685 BP; 206 A; 157 C; 138 G; 184 T; 0 other;

Query Match 79.0%; Score 15.8; DB 21; Length 685;
Best Local Similarity 89.5%; Pred. No. 32;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy      2 tcacgtgacacagtcaaac 20
          ||||  ||||||||||||||
Db     65 tcacccgacacagtcaaac 83


RESULT  11
AAZ57158
ID   AAZ57158 standard; cDNA; 738 BP.
XX
AC   AAZ57158;
XX
DT   03-APR-2000  (first entry)
XX
DE   Human type IV collagen alpha 3 chain nucleotide sequence SEQ ID NO:9.
XX
KW   Human; type IV collagen; anti-angiogenic; angiogenesis; cancer;
KW   benign tumour; rheumatoid arthritis; diabetic retinopathy; psoriasis;
KW   ocular angiogenesis disease; Osler-Webber Syndrome; telangiectasia;
KW   myocardial angiogenesis; plaque neovascularisation; angiofibroma;
KW   atherosclerosis; scleroderma; hypertrophic scar; cat scratch disease;
KW   contraception; obesity; ss.
XX
OS   Homo sapiens.
XX
PN   WO9965940-A1.
XX
PD   23-DEC-1999.
XX
PF   17-JUN-1999;   99WO-US13737.
XX
PR   17-JUN-1998;   98US-0089689.
PR   25-MAR-1999;   99US-0126175.
XX
PA   (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
XX
PI   Kalluri R;
XX
DR   WPI; 2000-097708/08.
DR   P-PSDB; AAY67942.
XX
PT   Anti-angiogenic proteins comprising the NC1 domain of the alpha 1, 2 or
PT   3 chain of Type IV collagen used in, e.g. treatment of benign tumors
PT   and rheumatoid arthritis -
XX
PS   Claim 33; Fig 16A; 117pp; English.
XX
CC   The present sequence encodes the human type IV collagen alpha 3 chain.
CC   The present invention describes an isolated protein chosen from the NC1
CC   domain of the alpha 1, alpha 2 or alpha 3 chains of type IV collagen or
CC   a fragment, analogue, derivative or mutant, which has anti-angiogenic
CC   properties. The anti-angiogenic proteins, multimers and chimeras are
CC   useful for inhibiting angiogenic activity in mammalian tissue,
CC   especially for treating diseases chosen from angiogenesis-dependent
CC   cancers, benign tumours, rheumatoid arthritis, diabetic retinopathy,
CC   psoriasis, ocular angiogenesis diseases, Osler-Webber Syndrome,
CC   myocardial angiogenesis, plaque neovascularisation, telangiectasia,
CC   haemophiliac joints, angiofibroma, wound granulation, intestinal
CC   adhesions, atherosclerosis, scleroderma, hypertrophic scars, cat scratch
CC   disease, Helicobacter pylori ulcers, dialysis graft vascular access
CC   stenosis, contraception and obesity. The compositions can be used to
CC   inhibit a disease characterised by angiogenic activity, in conjunction
CC   with radiation therapy, chemotherapy or immunotherapy.
XX
SQ   Sequence 738 BP; 204 A; 184 C; 162 G; 188 T; 0 other;


  Query Match             79.0%;  Score 15.8;  DB 21;  Length 738;
  Best Local Similarity   89.5%;  Pred. No. 32;
  Matches   17;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      2 tcacgtgacacagtcaaac 20
          ||||  ||||||||||||||
Db     68 tcacccgacacagtcaaac 86


RESULT  12
AAZ20091
ID   AAZ20091 standard; DNA; 900 BP.
XX
AC   AAZ20091;
XX
DT   05-JAN-2000  (first entry)
XX
DE   Type IV collagen NC1 domain alpha-3 monomer DNA.
XX
KW   Type IV collagen; NC1 domain; non-collagenous domain; human;
KW   angiogenesis; tumour; metastasis; therapy; diabetic retinopathy;
KW   rheumatoid arthritis; retinal neovascularization;
KW   choroidal neovascularization; macular degeneration;
KW   corneal neovascularization; retinopathy of prematurity;
KW   corneal graft rejection; neovascular glaucoma;
KW   retrolental fibroplasia; epidemic keratoconjunctivitis;
KW   vitamin A deficiency; contact lens overwear; atopic keratitis;
KW   superior limbic keratitis; pterygium keratitis sicca; sogrens;
KW   acne rosacea; phylectenulosis; syphilis; Mycobacteria infection;
KW   lipid degeneration; chemical burn; ulcer; herpes simplex infection;
KW   Herpes zoster infection; protozoan infection; Kaposi's sarcoma;
KW   Mooren ulcer; Terrien's marginal degeneration;
KW   marginal keratolysis; trauma; systemic lupus; polyarteritis;
KW   Wegener's sarcoidosis; scleritis; Steven's Johnson disease;
KW   radial keratotomy; sickle cell anaemia; sarcoid;
KW   pseudoxanthoma elasticum; Paget's disease; vein occlusion;
KW   artery occlusion; carotid obstructive disease; chronic uveitis;
KW   chronic vitritis; Lyme's disease; Eales disease; Bechets disease;
KW   myopia; optic pit; Stargart's disease; pars planitis;
KW   chronic retinal detachment; hyperviscosity syndrome; toxoplasmosis;
KW   post-laser complication; fibrovascular tissue proliferation;
KW   haemangioma; Osler-Weber-Rendu; AIDS; ocular neovascular disease;
KW   osteoarthritis; chronic inflammation; Crohn's disease;
KW   ulceritive colitis; psoriasis; atherosclerosis; pemphigoid; ss.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   CDS             40..846
FT                   /*tag=  a
FT   sig_peptide     40..90
FT                   /*tag=  b
FT                   /product= "BM40 signal peptide"
FT   mat_peptide     91..843
FT                   /*tag=  c
FT                   /product= "affinity-tagged alpha-3 monomer"
XX
PN   WO9949885-A2.
XX
PD   07-OCT-1999.
XX
PF   26-MAR-1999;   99WO-US06445.
XX
PR   27-MAR-1998;   98US-0079783.
PR   29-OCT-1998;   98US-0106170.
XX
PA   (UNIV ) UNIV KANSAS MEDICAL CENT.
XX
PI   Hudson BG,  Sarras MP;
XX
DR   WPI; 1999-601297/51.
DR   P-PSDB; AAY31993.
XX
PT   Inhibition of angiogenesis with non-collagenous alpha chain monomer
PT   useful for treating e.g. tumor growth or metastasis,
PT   neovascularisation, etc -
XX
PS   Disclosure; Fig 17c; 56pp; English.
```

XX
CC This is the nucleotide sequence of a recombinant DNA encoding a
CC type IV collagen non-collagenous (NC1) domain alpha-3 polypeptide
CC (see AAY31993) composed of a BM40 signal sequence (which is cleaved
CC from the mature protein) to facilitate protein secretion, and a
CC mature protein comprising an affinity tag (facilitates purification
CC and identification of the material) and the alpha-1 chain monomer.
CC The invention provides methods and kits for inhibiting angiogenesis,
CC tumour growth and metastasis, and endothelial cell interaction with
CC the extracellular matrix, each method comprising contacting the
CC tumour or animal tissue with 1 or more isolated type IV collagen
CC NC1 alpha chain monomer(s) selected from the group consisting of
CC alpha-1, alpha-2, alpha-3 and alpha-6 NC1 chain monomers. The
CC monomers can be produced via recombinant protein expression. The
CC polynucleotides and polypeptides are used to treat an angiogenesis-
CC mediated disorder or condition, especially selected from solid and
CC blood-borne tumours, diabetic retinopathy, rheumatoid arthritis,
CC retinal neovascularization, choroidal neovascularization, macular
CC degeneration, corneal neovascularization, retinopathy of prematurity,
CC corneal graft rejection, neovascular glaucoma, retrolental
CC fibroplasia, epidemic keratoconjunctivitis, vitamin A deficiency,
CC contact lens overwear, atopic keratitis, superior limbic keratitis,
CC pterygium keratitis sicca, sogrens, acne rosacea, phylectenulosis,
CC syphilis, mycobacteria infections, lipid degeneration, chemical
CC burns, bacterial ulcers, fungal ulcers, herpes simplex infections,
CC herpes zoster infections, protozoan infections, Kaposi's sarcoma,
CC Mooren ulcer, Terrien's marginal degeneration, marginal keratolysis,
CC trauma, systemic lupus, polyarteritis, Wegener's sarcoidosis,
CC scleritis, Steven's Johnson disease, radial keratotomy, sickle cell
CC anaemia, sarcoid, pseudoxanthoma elasticum, Pagets disease, vein
CC occlusion, artery occlusion, carotid obstructive disease, chronic
CC uveitis, chronic vitritis, Lyme's disease, Eales disease, Bechets
CC disease, myopia, optic pits, Stargarts disease, pars planitis,
CC chronic retinal detachment, hyperviscosity syndromes, toxoplasmosis,
CC post-laser complications, abnormal proliferation of fibrovascular
CC tissue, haemangiomas, Osler-Weber-Rendu, AIDS, ocular neovascular
CC disease, osteoarthritis, chronic inflammation, Crohn's disease,
CC ulceritive colitis, psoriasis, atherosclerosis, and pemphigoid (all
CC claimed).
XX
SQ Sequence 900 BP; 228 A; 243 C; 206 G; 223 T; 0 other;

Query Match 79.0%; Score 15.8; DB 20; Length 900;
Best Local Similarity 89.5%; Pred. No. 34;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 tcacgtgacacagtcaaac 20
|||| ||||||||||||
Db 176 tcacccgacacagtcaaac 194

RESULT 13
AAA90993
ID AAA90993 standard; DNA; 900 BP.
XX
AC AAA90993;
XX
DT 12-FEB-2001 (first entry)
XX
DE Human alpha3(IV)NC1 coding sequence.
XX
KW Type IV collagen alpha chain monomer; human; inhibitor; angiogenesis;
KW tumour growth; integrin receptor; carcinoma; sarcoma; rhabdomyosarcoma;
KW retinoblastoma; Ewing sarcoma; neuroblastoma; osteosarcoma; leukaemia;
KW diabetic retinopathy; rheumatoid arthritis; neovascularisation;
KW muscular degeneration; corneal graft rejection; vitamin A deficiency;
KW atopic keratitis; Mycobacteria infection; chemical burn; sarcoid;
KW Kaposi's sarcoma; sickle cell anaemia; carotid obstructive disease;
KW chronic inflammation; psoriasis; therapy; alpha3(IV)NC1; ds.
XX
OS Homo sapiens.
XX

FH Key Location/Qualifiers
FT CDS 40..846
FT /*tag= a
FT /product= alpha3(IV)NC1
XX
PN WO200059532-A1.
XX
PD 12-OCT-2000.
XX
PF 31-MAR-2000; 2000WO-US08678.
XX
PR 01-APR-1999; 99US-0127391.
XX
PA (BIOS-) BIOSTRATUM INC.
XX
PI Brooks P, Hudson B;
XX
DR WPI; 2000-664962/64.
DR P-PSDB; AAY97555.
XX
PT Use of antagonists of specific integrin receptors for inhibiting
PT angiogenesis, tumour growth or metastases, or endothelial cell
PT interactions with the extracellular matrix
XX
PS Disclosure; Fig 17c; 78pp; English.
XX
CC This sequence encodes a human type IV collagen alpha chain monomer,
CC designated alpha3(IV)NC1. The invention relates to a method for
CC inhibiting angiogenesis, tumour growth or metastases, or endothelial cell
CC interactions with the extracellular matrix, comprising contacting the
CC cells or tissue with a polypeptide composition containing antagonists of
CC specific integrin receptors. The methods and the antagonists are useful
CC for inhibiting angiogenesis, tumour growth or metastases, or endothelial
CC cell interaction with the extracellular matrix. The antagonists are also
CC useful for treating diseases and conditions with accompanying undesired
CC angiogenesis, e.g. solid and blood-borne tumours (e.g. melanomas,
CC carcinomas, sarcomas, rhabdomyosarcoma, retinoblastoma, Ewing sarcoma,
CC neuroblastoma, osteosarcoma or leukaemia). These are also applicable to
CC treating non-tumourigenic diseases and conditions with accompanying
CC undesired angiogenesis, e.g. diabetic retinopathy, rheumatoid arthritis,
CC retinal neovascularisation, choroidal neovascularisation, muscular
CC degeneration, corneal graft rejection, vitamin A deficiency, atopic
CC keratitis, Mycobacteria infections, chemical burns, Kaposi's sarcoma,
CC sickle cell anaemia, sarcoid, carotid obstructive disease, post-laser
CC complications, chronic inflammation or psoriasis.
XX
SQ Sequence 900 BP; 228 A; 243 C; 206 G; 223 T; 0 other;

Query Match 79.0%; Score 15.8; DB 21; Length 900;
Best Local Similarity 89.5%; Pred. No. 34;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 tcacgtgacacagtcaaac 20
|||| ||||||||||||
Db 176 tcacccgacacagtcaaac 194

RESULT 14
AAF13921/c
ID AAF13921 standard; cDNA; 1276 BP.
XX
AC AAF13921;
XX
DT 13-MAR-2001 (first entry)
XX
DE Aspergillus oryzae EST SEQ ID NO:6444.
XX
KW Multiple gene expression; filamentous fungal cell; EST;
KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;
KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;
KW culture condition; environmental stress; spore morphogenesis;
KW metabolic pathway engineering; catabolic pathway engineering; ss.

XX
OS Aspergillus oryzae.
XX
PN WO200056762-A2.
XX
PD 28-SEP-2000.
XX
PF 22-MAR-2000; 2000WO-US07781.
XX
PR 22-MAR-1999; 99US-0273623.
XX
PA (NOVO) NOVO NORDISK BIOTECH INC.
PA (NOVO) NOVO NORDISK AS.
XX
PI Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;
XX
DR WPI; 2000-594572/56.
XX
PT Monitoring differential expression of genes in filamentous fungal cells
PT uses fluorescence-labeled nucleic acids isolated from the cells and a
PT substrate of expressed sequence tags -
XX
PS Claim 88; Page 2643-2644; 3161pp; English.
XX
CC The present invention describes a method for monitoring differential
CC expression of genes in a first filamentous fungal (FF) cell relative to
CC expression of the same genes in one or more second filamentous fungal
CC cells. The method uses fluorescence-labeled nucleic acids isolated from
CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs
CC are used in the methods for monitoring differential expression of genes
CC in a first filamentous fungal (FF) cell relative to expression of the
CC same genes in one or more second filamentous fungal cells. Monitoring
CC the global expression of genes from FF cells allows the production
CC potential of the microorganisms to be improved. New genes may be
CC discovered, possible functions of unknown open reading frames can be
CC identified and gene copy number variation and stability can be
CC monitored. The expression of genes can be used to study how FF cells
CC adapt to changes in culture conditions, environmental stress, spore
CC morphogenesis, recombination, metabolic or catabolic pathway
CC engineering. Using ESTs provides several advantages over genomic or
CC random cDNA clones including elimination of redundancy as one spot on an
CC array equals one gene or open reading frame, and organisation of the
CC microarrays based on function of the gene products to facilitate
CC analysis of the results. AAF07478 to AAF11247 represents ESTs from
CC Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from Aspergillus
CC niger; AAF11854 to AAF14878 represents ESTs from Aspergillus oryzae; and
CC AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are
CC all specifically claimed in the present invention.
XX
SQ Sequence 1276 BP; 311 A; 319 C; 314 G; 332 T; 0 other;

Query Match 76.0%; Score 15.2; DB 21; Length 1276;
Best Local Similarity 85.0%; Pred. No. 73;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 gtcacgtgacacagtcaaac 20
||| |||||||||||| ||
Db 746 GTCTTGTGACACAGTCACAC 727

RESULT 15
AAC35997
ID AAC35997 standard; DNA; 1801 BP.
XX
AC AAC35997;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 12176.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.

PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 76.0%; Score 15.2; DB 21; Length 1801;
Best Local Similarity 85.0%; Pred. No. 78;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
Qy       1 gtcacgtgacacagtcaaac 20
           |||| |||| ||| ||||||
Db    1539 gtcaagtgagacactcaaac 1558
```

Search completed: July 25, 2001, 05:23:05
Job time: 4678 sec

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: July 25, 2001, 05:18:52 ; Search time 117.39 Seconds
(without alignments)
31.563 Million cell updates/sec

Title: US-09-142-095-3
Perfect score: 20
Sequence: 1 gtcacgtgacacagtcaaac 20

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 317530 seqs, 92630169 residues

Total number of hits satisfying chosen parameters: 635060

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	16.4	82.0	22	3	US-08-857-464-3	Sequence 3, Appli
c	2	14.8	74.0	519	4	US-08-991-789A-270	Sequence 270, App
	3	14.8	74.0	1990	2	US-08-967-101-154	Sequence 154, App
	4	14.8	74.0	1990	2	US-08-592-541-154	Sequence 154, App
	5	14.8	74.0	1990	4	US-09-124-698-154	Sequence 154, App
	6	14.8	74.0	1990	4	US-09-127-480-154	Sequence 154, App
	7	14.8	74.0	1990	4	US-08-496-841C-154	Sequence 154, App
	8	14.8	74.0	2740	1	US-08-100-247-4	Sequence 4, Appli
	9	14.8	74.0	2740	1	US-08-483-146A-4	Sequence 4, Appli
	10	14.8	74.0	2740	1	US-08-232-513A-5	Sequence 5, Appli
	11	14.8	74.0	2740	1	US-08-484-594A-4	Sequence 4, Appli
c	12	14.4	72.0	448	4	US-09-060-756-666	Sequence 666, App
c	13	14.2	71.0	84	1	US-08-053-131-145	Sequence 145, App
c	14	14.2	71.0	84	1	US-08-645-641-145	Sequence 145, App
c	15	14.2	71.0	84	1	US-07-853-408B-145	Sequence 145, App
c	16	14.2	71.0	84	1	US-08-096-762-145	Sequence 145, App
c	17	14.2	71.0	84	2	US-08-308-865-145	Sequence 145, App
c	18	14.2	71.0	84	5	PCT-US92-10983-145	Sequence 145, App
c	19	14.2	71.0	2700	3	US-09-022-875-1	Sequence 1, Appli
c	20	13.8	69.0	410	4	US-09-060-756-239	Sequence 239, App
c	21	13.8	69.0	453	4	US-09-060-756-189	Sequence 189, App
	22	13.8	69.0	1096	4	US-09-056-153-1	Sequence 1, Appli
	23	13.8	69.0	1941	4	US-09-402-002-1	Sequence 47, Appl
	24	13.8	69.0	2336	4	US-08-796-101-47	Sequence 215, App
	25	13.8	69.0	3022	4	US-08-961-083-215	Sequence 7, Appli
	26	13.8	69.0	4977	1	US-08-030-096-7	Sequence 49, Appl
c	27	13.8	69.0	5275	4	US-08-796-101-49	
c	28	13.8	69.0	35100	2	US-08-770-379-18	Sequence 18, Appl
c	29	13.8	69.0	35100	4	US-08-757-669A-18	Sequence 18, Appl
c	30	13.6	68.0	30	1	US-08-480-784-40	Sequence 40, Appl
c	31	13.6	68.0	30	1	US-08-483-553-40	Sequence 40, Appl
c	32	13.6	68.0	30	1	US-08-487-002-40	Sequence 40, Appl
c	33	13.6	68.0	30	1	US-08-483-554B-40	Sequence 40, Appl
c	34	13.6	68.0	30	1	US-08-488-011B-40	Sequence 40, Appl
c	35	13.6	68.0	30	4	US-08-850-727-40	Sequence 40, Appl
c	36	13.6	68.0	30	5	PCT-US95-10202-40	Sequence 40, Appl
c	37	13.6	68.0	30	5	PCT-US95-10203-40	Sequence 40, Appl
c	38	13.6	68.0	30	5	PCT-US95-10220-40	Sequence 40, Appl
	39	13.6	68.0	372	4	US-09-124-671-30	Sequence 30, Appl
c	40	13.6	68.0	631	1	US-08-480-784-15	Sequence 15, Appl
c	41	13.6	68.0	631	1	US-08-483-553-15	Sequence 15, Appl
c	42	13.6	68.0	631	1	US-08-487-002-15	Sequence 15, Appl
c	43	13.6	68.0	631	1	US-08-483-554B-15	Sequence 15, Appl
c	44	13.6	68.0	631	1	US-08-488-011B-15	Sequence 15, Appl
c	45	13.6	68.0	631	4	US-08-850-727-15	Sequence 15, Appl

ALIGNMENTS

RESULT 1
US-08-857-464-3/c
; Sequence 3, Application US/08857464
; Patent No. 6066450
; GENERAL INFORMATION:
; APPLICANT: Ralston, Stuart H.
; APPLICANT: Grant, Struan F.A.
; TITLE OF INVENTION: Diagnostic and Therapeutic Methods and Apparatus
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
; STREET: 1100 New York Avenue, N.W., Suite 600
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/857,464
; FILING DATE: 16-MAY-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9610281.9
; FILING DATE: 16-MAY-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Esmond, Robert W.
; REGISTRATION NUMBER: 32,893
; REFERENCE/DOCKET NUMBER: 1581.0180000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-2600
; TELEFAX: 202-371-2540
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 22 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: both
; MOLECULE TYPE: cDNA
US-08-857-464-3

Query Match 82.0%; Score 16.4; DB 3; Length 22;
Best Local Similarity 94.4%; Pred. No. 4.2;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 tcacgtgacacagtcaaa 19

```
                  ||||||||||||| |||||
Db      19 TCACGTGACACACTCAAA 2


RESULT   2
US-08-991-789A-270/c
; Sequence 270, Application US/08991789A
; Patent No. 6225054
;   GENERAL INFORMATION:
;        APPLICANT: Frudakis, Tony N.
;                   Smith, John M.
;                   Reed, Steven G.
;        TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
;                            TREATMENT AND DIAGNOSIS OF BREAST CANCER
;        NUMBER OF SEQUENCES: 292
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Seed IP Law Group
;             STREET: 701 Fifth Avenue, Suite 6300
;             CITY: Seattle
;             STATE: Washington
;             COUNTRY: USA
;             ZIP: 98104-7092
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/08/991,789A
;             FILING DATE: 11-Dec-1997
;             CLASSIFICATION: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Potter, Jane E. R.
;             REGISTRATION NUMBER: 33,332
;             REFERENCE/DOCKET NUMBER: 210121.419C3
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (206) 622-4900
;             TELEFAX: (206) 682-6031
;   INFORMATION FOR SEQ ID NO: 270:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 519 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        SEQUENCE DESCRIPTION: SEQ ID NO: 270:
US-08-991-789A-270


  Query Match           74.0%;  Score 14.8;  DB 4;  Length 519;
  Best Local Similarity  88.9%;  Pred. No. 41;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      2 tcacgtgacacagtcaaa 19
          ||||||||| ||| ||||
Db     74 TCACGTGACTCAGCCAAA 57


RESULT   3
US-08-967-101-154
; Sequence 154, Application US/08967101
; Patent No. 5840540
;  GENERAL INFORMATION:
;    APPLICANT:  ST. GEORGE-HYSLOP, PETER H
;    APPLICANT:  ROMMENS, JOHANNA M
;    APPLICANT:  FRASER, PAUL E
;    TITLE OF INVENTION:  GENETIC SEQUENCES AND PROTEINS RELATED
;    TITLE OF INVENTION:  TO ALZHEIMER'S DISEASE
;    NUMBER OF SEQUENCES:  183
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  TESTA, HURWITZ & THIBEAULT
;      STREET:  High Street Tower - 125 High Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  U.S.A.
;      ZIP:  02110
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/967,101
;      FILING DATE:  10-NOV-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/592,541
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Pitcher, Edmund R.
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 248-7000
;      TELEFAX:  (617) 248-7100
;  INFORMATION FOR SEQ ID NO:  154:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1990 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-967-101-154


  Query Match           74.0%;  Score 14.8;  DB 2;  Length 1990;
  Best Local Similarity  84.2%;  Pred. No. 49;
  Matches   16;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      2 tcacgtgacacagtcaaac 20
          ||| |||||| ||| |||||
Db    168 TCAAGTGACCCAGNCAAAC 186


RESULT   4
US-08-592-541-154
; Sequence 154, Application US/08592541
; Patent No. 5986054
;  GENERAL INFORMATION:
;    APPLICANT:  ST. GEORGE-HYSLOP, PETER H
;    APPLICANT:  ROMMENS, JOHANNA M
;    APPLICANT:  FRASER, PAUL E
;    TITLE OF INVENTION:  GENETIC SEQUENCES AND PROTEINS RELATED
;    TITLE OF INVENTION:  TO ALZHEIMER'S DISEASE
;    NUMBER OF SEQUENCES:  183
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  TESTA, HURWITZ & THIBEAULT
;      STREET:  High Street Tower - 125 High Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  U.S.A.
;      ZIP:  02110
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/592,541
;      FILING DATE:
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Pitcher, Edmund R.
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 248-7000
;      TELEFAX:  (617) 248-7100
;  INFORMATION FOR SEQ ID NO:  154:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1990 base pairs
```

```
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-592-541-154


  Query Match             74.0%;  Score 14.8;  DB 2;  Length 1990;
  Best Local Similarity   84.2%;  Pred. No. 49;
  Matches   16;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 tcacgtgacacagtcaaac 20
           ||| ||||| ||| |||||
Db     168 TCAAGTGACCCAGNCAAAC 186


RESULT   5
US-09-124-698-154
; Sequence 154, Application US/09124698
; Patent No. 6117978
;  GENERAL INFORMATION:
;    APPLICANT:  ST. GEORGE-HYSLOP, PETER H
;    APPLICANT:  ROMMENS, JOHANNA M
;    APPLICANT:  FRASER, PAUL E
;    TITLE OF INVENTION:  GENETIC SEQUENCES AND PROTEINS RELATED
;    TITLE OF INVENTION:  TO ALZHEIMER'S DISEASE
;    NUMBER OF SEQUENCES:  183
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  TESTA, HURWITZ & THIBEAULT
;      STREET:  High Street Tower - 125 High Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  U.S.A.
;      ZIP:  02110
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/124,698
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/592,541
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Pitcher, Edmund R.
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 248-7000
;      TELEFAX:  (617) 248-7100
;  INFORMATION FOR SEQ ID NO:  154:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1990 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-09-124-698-154


  Query Match             74.0%;  Score 14.8;  DB 4;  Length 1990;
  Best Local Similarity   84.2%;  Pred. No. 49;
  Matches   16;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 tcacgtgacacagtcaaac 20
           ||| ||||| ||| |||||
Db     168 TCAAGTGACCCAGNCAAAC 186


RESULT   6
US-09-127-480-154
; Sequence 154, Application US/09127480
; Patent No. 6194153
;  GENERAL INFORMATION:
;    APPLICANT:  ST. GEORGE-HYSLOP, PETER H
;    APPLICANT:  ROMMENS, JOHANNA M
;    APPLICANT:  FRASER, PAUL E
;    TITLE OF INVENTION:  GENETIC SEQUENCES AND PROTEINS RELATED
;    TITLE OF INVENTION:  TO ALZHEIMER'S DISEASE
;    NUMBER OF SEQUENCES:  183
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  TESTA, HURWITZ & THIBEAULT
;      STREET:  High Street Tower - 125 High Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  U.S.A.
;      ZIP:  02110
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/127,480
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/592,541
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Pitcher, Edmund R.
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 248-7000
;      TELEFAX:  (617) 248-7100
;  INFORMATION FOR SEQ ID NO:  154:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1990 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-09-127-480-154


  Query Match             74.0%;  Score 14.8;  DB 4;  Length 1990;
  Best Local Similarity   84.2%;  Pred. No. 49;
  Matches   16;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 tcacgtgacacagtcaaac 20
           ||| ||||| ||| |||||
Db     168 TCAAGTGACCCAGNCAAAC 186


RESULT   7
US-08-496-841C-154
; Sequence 154, Application US/08496841C
; Patent No. 6210919
;  GENERAL INFORMATION:
;      APPLICANT: ST. GEORGE-HYSLOP, PETER H
;                 ROMMENS, JOHANNA M
;                 FRASER, PAUL E
;      TITLE OF INVENTION: GENETIC SEQUENCES AND PROTEINS RELATED
;                          TO ALZHEIMER'S DISEASE
;      NUMBER OF SEQUENCES: 175
;      CORRESPONDENCE ADDRESS:
;           ADDRESSEE: Darby & Darby, PC
;           STREET: 805 Third Avenue
;           CITY: New York
;           STATE: New York
;           COUNTRY: U.S.A.
;           ZIP: 10022
;      COMPUTER READABLE FORM:
;           MEDIUM TYPE: Floppy disk
;           COMPUTER: IBM PC compatible
;           OPERATING SYSTEM: PC-DOS/MS-DOS
```

```
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/08/496,841C
;            FILING DATE: 28-Jun-1995
;            CLASSIFICATION: <Unknown>
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Paul F. Fehlner, Ph.D.
;            REGISTRATION NUMBER: 35,135
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (212) 527-7700
;            TELEFAX: (212) 753-6237
;   INFORMATION FOR SEQ ID NO: 154:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 1990 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: single
;            TOPOLOGY: linear
;       MOLECULE TYPE: cDNA
;       SEQUENCE DESCRIPTION: SEQ ID NO: 154:
US-08-496-841C-154


  Query Match             74.0%;  Score 14.8;  DB 4;  Length 1990;
  Best Local Similarity   84.2%;  Pred. No. 49;
  Matches   16;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       2 tcacgtgacacagtcaaac 20
           ||| ||||| ||| |||||
Db     168 TCAAGTGACCCAGNCAAAC 186


RESULT   8
US-08-100-247-4
; Sequence 4, Application US/08100247
; Patent No. 5571787
;  GENERAL INFORMATION:
;    APPLICANT:  O'BRIEN, JOHN S.
;    APPLICANT:  KISHIMOTO, YASUO
;    TITLE OF INVENTION:  PROSAPOSIN AS A NEUROTROPHIC FACTOR
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  KNOBBE, MARTENS, OLSON AND BEAR
;      STREET:  620 NEWPORT CENTER DRIVE   SIXTEENTH FLOOR
;      CITY:  NEWPORT BEACH
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  92660
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/100,247
;      FILING DATE:  19930730
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Israelsen, Ned A.
;      REGISTRATION NUMBER:  29,655
;      REFERENCE/DOCKET NUMBER:  OBRIEN.002A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  619-235-8550
;      TELEFAX:  619-235-0176
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2740 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    IMMEDIATE SOURCE:
;      CLONE:  PROSAPOSIN cDNA
US-08-100-247-4


  Query Match             74.0%;  Score 14.8;  DB 1;  Length 2740;
  Best Local Similarity   88.9%;  Pred. No. 51;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 tcacgtgacacagtcaaa 19
           |||||||||| ||| ||||
Db    1194 TCACGTGACTCAGCCAAA 1211


RESULT   9
US-08-483-146A-4
; Sequence 4, Application US/08483146A
; Patent No. 5696080
;  GENERAL INFORMATION:
;    APPLICANT:  O'Brien, John S.
;    APPLICANT:  Kishimoto, Yasuo
;    TITLE OF INVENTION:  PHARMACEUTICAL COMPOSITIONS
;    TITLE OF INVENTION:  COMPRISING PROSAPOSIN AND NEUROTROPHIC PEPTIDES DERIVED
;    TITLE OF INVENTION:  THEREFROM
;    NUMBER OF SEQUENCES:  11
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Knobbe, Martens, Olson and Bear
;      STREET:  620 Newport Center Blvd.  16th Floor
;      CITY:  Newport Beach
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  92660
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/483,146A
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Israelsen, Ned A
;      REGISTRATION NUMBER:  29,655
;      REFERENCE/DOCKET NUMBER:  MYELOS.002DV1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  619-235-8550
;      TELEFAX:  619-235-0176
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2740 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-483-146A-4


  Query Match             74.0%;  Score 14.8;  DB 1;  Length 2740;
  Best Local Similarity   88.9%;  Pred. No. 51;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 tcacgtgacacagtcaaa 19
           |||||||||| ||| ||||
Db    1194 TCACGTGACTCAGCCAAA 1211


RESULT  10
US-08-232-513A-5
; Sequence 5, Application US/08232513A
```

```
; Patent No. 5700909
;  GENERAL INFORMATION:
;    APPLICANT:  O'Brien, John S.
;    TITLE OF INVENTION:  Prosaposin and Cytokine-Derived Peptides
;    TITLE OF INVENTION:  as Therapeutic Agents
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Campbell & Flores LLP
;      STREET:  4370 La Jolla Village Drive, Suite 700
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  92122
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/232,513A
;      FILING DATE:  21-APR-1994
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/100,247
;      FILING DATE:  30-JUL-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Campbell, Cathryn A.
;      REGISTRATION NUMBER:  31,815
;      REFERENCE/DOCKET NUMBER:  P-UD 1643
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 535-9001
;      TELEFAX:  (619) 535-8949
;   INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2740 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1..2740
;      OTHER INFORMATION:  /label= Hum_prosaposin
US-08-232-513A-5


  Query Match             74.0%;  Score 14.8;  DB 1;  Length 2740;
  Best Local Similarity   88.9%;  Pred. No. 51;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      2 tcacgtgacacagtcaaa 19
          ||||||||| ||| ||||
Db   1194 TCACGTGACTCAGCCAAA 1211


RESULT  11
US-08-484-594A-4
; Sequence 4, Application US/08484594A
; Patent No. 5714459
;  GENERAL INFORMATION:
;    APPLICANT:  O'Brien, John S.
;    APPLICANT:  Kishimoto, Yasuo
;    TITLE OF INVENTION:  USE OF PROSAPOSIN AND NEUROTROPHIC PEPTIDES
;    TITLE OF INVENTION:  DERIVED THEREFROM
;    NUMBER OF SEQUENCES:  11
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Knobbe, Martens, Olson and Bear
;      STREET:  620 Newport Center Drive, Sixteenth Floor
;      CITY:  Newport Beach
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  92660
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/484,594A
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/100,247
;      FILING DATE:  30-JUL-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Israelsen, Ned A
;      REGISTRATION NUMBER:  29,655
;      REFERENCE/DOCKET NUMBER:  MYELOS.002DV2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  619-235-8550
;      TELEFAX:  619-235-0176
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2740 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-484-594A-4


  Query Match             74.0%;  Score 14.8;  DB 1;  Length 2740;
  Best Local Similarity   88.9%;  Pred. No. 51;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      2 tcacgtgacacagtcaaa 19
          ||||||||| ||| ||||
Db   1194 TCACGTGACTCAGCCAAA 1211


RESULT  12
US-09-060-756-666/c
; Sequence 666, Application US/09060756
; Patent No. 6183957
; GENERAL INFORMATION:
;  APPLICANT: Cole, Stewart
;  APPLICANT: Buchrieser-Brosch, Roland
;  APPLICANT: Gordon, Stephen
;  APPLICANT: Billault, Alain
;  TITLE OF INVENTION: METHOD FOR ISOLATING A POLYNUCLEOTIDE OF INTEREST FROM
;  TITLE OF INVENTION: THE GENOME OF A MYCOBACTERIUM USING A BAC-BASED DNA
;  TITLE OF INVENTION: LIBRARY APPLICATION TO THE DETECTION OF MYCOBACTERIA
;  FILE REFERENCE: 3495-0169
;  CURRENT APPLICATION NUMBER: US/09/060,756
;  CURRENT FILING DATE: 1998-04-16
;  NUMBER OF SEQ ID NOS: 743
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 666
;   LENGTH: 448
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (various positions within the sequence)
;   OTHER INFORMATION: applicants are uncertain of bases designated as "n"
US-09-060-756-666


  Query Match             72.0%;  Score 14.4;  DB 4;  Length 448;
  Best Local Similarity   88.2%;  Pred. No. 65;
  Matches   15;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      4 acgtgacacagtcaaac 20
          ||| ||||| |||||||
Db    298 ACGGGACACNGTCAAAC 282
```

```
RESULT 13
US-08-053-131-145/c
; Sequence 145, Application US/08053131
; Patent No. 5661016
;  GENERAL INFORMATION:
;    APPLICANT:  Lonberg, Nils
;    APPLICANT:  Kay, Robert M.
;    TITLE OF INVENTION:  Transgenic No. 5661016-Human Animals for
;    TITLE OF INVENTION:  Producing Heterologous Antibodies
;    NUMBER OF SEQUENCES:  197
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Townsend and Townsend Khourie and Crew
;      STREET:  One Market Plaza, Steuart Tower, Suite 200
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94105
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/053,131
;      FILING DATE:  26-APR-1993
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/990,860
;      FILING DATE:  16-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/810,279
;      FILING DATE:  17-DEC-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/853,408
;      FILING DATE:  18-MAR-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Smith, William M.
;      REGISTRATION NUMBER:  30,223
;      REFERENCE/DOCKET NUMBER:  14643-9-3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-326-2400
;      TELEFAX:  415-326-2422
;  INFORMATION FOR SEQ ID NO:  145:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  84 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-053-131-145


  Query Match           71.0%;  Score 14.2;  DB 1;  Length 84;
  Best Local Similarity  84.2%;  Pred. No. 66;
  Matches   16;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 gtcacgtgacacagtcaaa 19
           ||||||| |  ||||||||
Db      36 GTCACGTTAACCAGTCAAA 18


RESULT 14
US-08-645-641-145/c
; Sequence 145, Application US/08645641
; Patent No. 5719032
;  GENERAL INFORMATION:
;    APPLICANT:  Lonberg, Nils
;    APPLICANT:  Kay, Robert M.
;    TITLE OF INVENTION:  Transgenic No. 5719032-Human Animals for
;    TITLE OF INVENTION:  Producing Heterologous Antibodies
;    NUMBER OF SEQUENCES:  150
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  William M. Smith
;      STREET:  Two Embarcadero Center, 8th Floor
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94111-3834
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/645,641
;      FILING DATE:  20-MAY-1996
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/904,068
;      FILING DATE:  23-JUN-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Smith, William M.
;      REGISTRATION NUMBER:  30,223
;      REFERENCE/DOCKET NUMBER:  14643-000913
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-326-2400
;      TELEFAX:  415-326-2422
;  INFORMATION FOR SEQ ID NO:  145:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  84 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-645-641-145


  Query Match           71.0%;  Score 14.2;  DB 1;  Length 84;
  Best Local Similarity  84.2%;  Pred. No. 66;
  Matches   16;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 gtcacgtgacacagtcaaa 19
           ||||||| |  ||||||||
Db      36 GTCACGTTAACCAGTCAAA 18


RESULT 15
US-07-853-408B-145/c
; Sequence 145, Application US/07853408B
; Patent No. 5789650
;  GENERAL INFORMATION:
;    APPLICANT:  Lonberg, Nils
;    APPLICANT:  Kay, Robert M.
;    TITLE OF INVENTION:  Transgenic No. 5789650-Human Animals for
;    TITLE OF INVENTION:  Producing Heterologous Antibodies
;    NUMBER OF SEQUENCES:  150
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  William M. Smith
;      STREET:  One Market Plaza, Steuart Tower, Suite 2000
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94105
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/853,408B
;      FILING DATE:  19920318
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Smith, William M.
```

; REGISTRATION NUMBER: 30,223
; REFERENCE/DOCKET NUMBER: 14643-9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-326-2400
; TELEFAX: 415-326-2422
; INFORMATION FOR SEQ ID NO: 145:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 84 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-07-853-408B-145

Query Match 71.0%; Score 14.2; DB 1; Length 84;
Best Local Similarity 84.2%; Pred. No. 66;
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
Qy      1 gtcacgtgacacagtcaaa 19
          ||||||| |  ||||||||
Db     36 GTCACGTTAACCAGTCAAA 18
```

Search completed: July 25, 2001, 05:18:53
Job time: 9216 sec

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: July 25, 2001, 04:54:51 ; Search time 2762.24 Seconds
(without alignments)
68.443 Million cell updates/sec

Title: US-09-142-095-3
Perfect score: 20
Sequence: 1 gtcacgtgacacagtcaaac 20

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est33:*
26: gb_est34:*
27: gb_est35:*
28: gb_est36:*
29: gb_est37:*
30: gb_est38:*
31: gb_est39:*
32: gb_est40:*
33: em_estba:*
34: em_estfun:*
35: em_esthum1:*
36: em_esthum2:*
37: em_esthum3:*
38: em_esthum4:*
39: em_esthum5:*
40: em_esthum6:*
41: em_esthum7:*
42: em_esthum8:*
43: em_esthum9:*
44: em_esthum10:*
45: em_esthum11:*
46: em_esthum12:*
47: em_esthum13:*
48: em_esthum14:*
49: em_esthum15:*
50: em_esthum16:*
51: em_esthum17:*
52: em_esthum18:*
53: em_esthum19:*
54: em_esthum20:*
55: em_esthum21:*
56: em_esthum22:*
57: em_esthum23:*
58: em_esthum24:*
59: em_esthum25:*
60: em_esthum26:*
61: em_esthum27:*
62: em_esthum28:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estin5:*
68: em_estom1:*
69: em_estom2:*
70: em_estov1:*
71: em_estov2:*
72: em_estpl1:*
73: em_estpl2:*
74: em_estpl3:*
75: em_estpl4:*
76: em_estpl5:*
77: em_estpl6:*
78: em_estpl7:*
79: em_estpl8:*
80: em_estpl9:*
81: em_estpl10:*
82: em_estro1:*
83: em_estro2:*
84: em_estro3:*
85: em_estro4:*
86: em_estro5:*
87: em_estro6:*
88: em_estro7:*
89: em_estro8:*
90: em_estro9:*
91: em_estro10:*
92: em_estro11:*
93: em_estro12:*
94: em_estro13:*
95: em_estro14:*
96: em_estro15:*
97: em_estro16:*
98: em_estro17:*
99: em_estro18:*
100: em_estro19:*
101: em_estro20:*
102: gb_est25:*
103: gb_est26:*
104: gb_est27:*
105: gb_est28:*
106: gb_est29:*
107: gb_est30:*
108: gb_est31:*
109: gb_est32:*
110: gb_est41:*
111: gb_est42:*
112: gb_est43:*
113: gb_est44:*
114: gb_est45:*
115: gb_est46:*
116: gb_est47:*

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est75:*
137: gb_est76:*
138: gb_est77:*
139: gb_est78:*
140: gb_est79:*
141: gb_est80:*
142: gb_est81:*
143: gb_est82:*
144: gb_est83:*
145: gb_est84:*
146: gb_est85:*
147: gb_est86:*
148: gb_est87:*
149: gb_est88:*
150: gb_est89:*
151: gb_est90:*
152: gb_est99:*
153: gb_est100:*
154: gb_est101:*
155: gb_est102:*
156: gb_est103:*
157: gb_est104:*
158: gb_est105:*
159: gb_est106:*
160: gb_est67:*
161: gb_est68:*
162: gb_est69:*
163: gb_est70:*
164: gb_est71:*
165: gb_est72:*
166: gb_est73:*
167: gb_est74:*
168: gb_est91:*
169: gb_est92:*
170: gb_est93:*
171: gb_est94:*
172: gb_est95:*
173: gb_est96:*
174: gb_est97:*
175: gb_est98:*
176: em_esthum29:*
177: em_esthum30:*
178: em_esthum31:*
179: em_esthum32:*
180: em_esthum33:*
181: em_estom3:*
182: em_estpl11:*
183: em_estro21:*
184: em_estro22:*
185: em_estro23:*
186: em_htc:*
187: gb_est107:*
188: gb_est108:*
189: gb_est109:*
190: gb_est110:*
191: gb_est111:*
192: gb_htc:*
193: em_gss_fun:*
194: em_gss_hum1:*
195: em_gss_hum2:*
196: em_gss_hum3:*
197: em_gss_hum4:*
198: em_gss_hum5:*
199: em_gss_hum6:*
200: em_gss_hum7:*
201: em_gss_hum8:*
202: em_gss_hum9:*
203: em_gss_inv1:*
204: em_gss_inv2:*
205: em_gss_inv3:*
206: em_gss_other:*
207: em_gss_pln1:*
208: em_gss_pln2:*
209: em_gss_pro:*
210: em_gss_rod1:*
211: em_gss_rod2:*
212: em_gss_rod3:*
213: em_gss_rod4:*
214: em_gss_rod5:*
215: em_gss_vrt1:*
216: em_gss_vrt2:*
217: em_gss_vrt3:*
218: gb_gss1:*
219: gb_gss2:*
220: gb_gss3:*
221: gb_gss4:*
222: gb_gss5:*
223: gb_gss6:*
224: gb_gss7:*
225: gb_gss8:*
226: gb_gss9:*
227: gb_gss10:*
228: gb_gss11:*
229: gb_gss12:*
230: gb_gss13:*
231: gb_gss14:*
232: gb_gss15:*
233: gb_gss16:*
234: gb_gss17:*
235: gb_gss18:*
236: gb_gss19:*
237: gb_gss20:*
238: gb_gss21:*
239: gb_gss22:*
240: gb_gss23:*
241: gb_gss24:*
242: gb_gss25:*
243: gb_gss26:*
244: gb_gss27:*
245: gb_gss28:*
246: gb_gss29:*
247: gb_gss30:*
248: gb_gss31:*
249: gb_gss32:*
250: gb_gss33:*
251: gb_gss34:*
252: em_gss_inv4:*
253: em_gss_rod6:*
254: em_gss_rod7:*
255: em_gss_rod8:*
256: gb_gss35:*
257: gb_gss36:*
258: gb_gss37:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
c 1	18.4	92.0	737	166	BE306792	BE306792 601101950
c 2	17.4	87.0	238	130	BB310653	BB310653 BB310653
3	17.4	87.0	295	16	AI104756	AI104756 EST214045
4	17.4	87.0	438	237	AZ015172	AZ015172 RPCI-23-2
5	17.4	87.0	482	233	AQ798101	AQ798101 HS_3158_A
c 6	17.4	87.0	554	17	AI180191	AI180191 EST223932
c 7	16.8	84.0	258	222	FR0039113	AL126611 Fugu rubr
c 8	16.8	84.0	400	222	FR0039183	AL126681 Fugu rubr
c 9	16.8	84.0	513	222	FR0039084	AL126582 Fugu rubr
10	16.8	84.0	539	166	BE368176	BE368176 601222089
c 11	16.8	84.0	548	149	BF500440	BF500440 AT15369.5
c 12	16.8	84.0	562	190	W73212	W73212 zd53e10.r1
c 13	16.8	84.0	804	146	BF232638	BF232638 602023221
c 14	16.8	84.0	835	220	CNS02G0P	AL195730 Tetraodon
15	16.8	84.0	835	221	CNS03EDV	AL240268 Tetraodon
16	16.8	84.0	968	220	CNS02RA3	AL210324 Tetraodon
17	16.8	84.0	1046	172	BF982699	BF982699 602305085
18	16.8	84.0	1114	154	BG493528	BG493528 602542446
19	16.4	82.0	597	231	AQ657757	AQ657757 Sheared D
c 20	16.4	82.0	619	222	FR0007351	Z91161 F.rubripes
21	16.4	82.0	837	221	CNS03PQN	AL254984 Tetraodon
22	16.4	82.0	873	150	BF578047	BF578047 602091895
c 23	16.4	82.0	892	220	CNS020LQ	AL175751 Tetraodon
c 24	16.4	82.0	988	219	CNS01PDJ	AL154679 Anopheles
c 25	16	80.0	971	221	CNS04LFN	AL296060 Tetraodon
c 26	16	80.0	981	221	CNS044PA	AL274375 Tetraodon
27	15.8	79.0	181	250	AZ841578	AZ841578 2M0139F09
28	15.8	79.0	199	127	BB160957	BB160957 BB160957
29	15.8	79.0	253	152	BG318736	BG318736 NXPV_016_
c 30	15.8	79.0	264	29	AV329412	AV329412 AV329412
c 31	15.8	79.0	279	109	AV062353	AV062353 AV062353
c 32	15.8	79.0	284	25	AV158238	AV158238 AV158238
c 33	15.8	79.0	287	16	AI111900	AI111900 UI-R-A1-e
34	15.8	79.0	303	14	AA957109	AA957109 UI-R-E1-f
c 35	15.8	79.0	311	15	AI040776	AI040776 ox15h08.s
c 36	15.8	79.0	327	17	AI228082	AI228082 EST224777
c 37	15.8	79.0	345	22	AI599772	AI599772 EST251475
c 38	15.8	79.0	370	224	AQ130228	AQ130228 HS_3007_A
c 39	15.8	79.0	375	112	AW142511	AW142511 EST292763
c 40	15.8	79.0	381	114	AW295307	AW295307 UI-H-BI2-
c 41	15.8	79.0	387	122	AW919078	AW919078 EST350382
c 42	15.8	79.0	400	188	R92042	R92042 yp96a04.r1
43	15.8	79.0	404	187	R17651	R17651 yg15g08.r1
44	15.8	79.0	409	14	AA944486	AA944486 EST199985
45	15.8	79.0	420	12	AA800910	AA800910 EST190407

ALIGNMENTS

```
RESULT  1
BE306792/c
LOCUS       BE306792      737 bp    mRNA            EST       26-OCT-2000
DEFINITION  601101950F1 NCI_CGAP_Lu29 Mus musculus cDNA clone IMAGE:3494340 5',
            mRNA sequence.
ACCESSION   BE306792
VERSION     BE306792.1  GI:9160544
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 737)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM8543  row: c  column: 13
            High quality sequence stop: 359.
FEATURES             Location/Qualifiers
     source          1. .737
                     /organism="Mus musculus"
                     /strain="CZECH II (feral)"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:3494340"
                     /clone_lib="NCI_CGAP_Lu29"
                     /tissue_type="spontaneous tumor, metastatic to mammary.
                     Stem cell origin."
                     /lab_host="DH10B"
                     /note="Organ: lung; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                     Library constructed by Life Technologies. Investigator
                     providing samples: Gilbert Smith, NIH"
BASE COUNT      169 a    190 c    177 g    201 t
ORIGIN


  Query Match           92.0%;  Score 18.4;  DB 166;  Length 737;
  Best Local Similarity  95.0%;  Pred. No. 31;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Caps    0;

Qy       1 gtcacgtgacacagtcaaac 20
           |||||||||||||||||| ||
Db     305 GTCACGTGACACAGTCACAC 286


RESULT  2
BB310653/c
LOCUS       BB310653      238 bp    mRNA            EST       10-JUL-2000
DEFINITION  BB310653 RIKEN full-length enriched, adult male corpora
            quadrigemina Mus musculus cDNA clone B230316K05 3', mRNA sequence.
ACCESSION   BB310653
VERSION     BB310653.1  GI:9011358
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 238)
  AUTHORS   Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci
            ,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N.,
            Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M.,
            Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N.,
            Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M.
            , Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.
            , Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata
            ,Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y.
            , Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tominaga,N., Toya
            ,T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaka,I.
            , Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino
            ,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Konno,H., et al.)
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Yoshihide Hayashizaki
            Genome Exploration Research Group, Life Science Tsukuba Center,
            Genome Science Laboratory
            The Institute of Physical and Chemical Research (RIKEN), Genomic
            Sciences Center
            3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
            Tel: +81-298-36-9013
            Fax: +81-298-36-9098
            Email: genome-res@rtc.riken.go.jp,
```

URL:http://genome.rtc.riken.go.jp/
Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki ,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Thermostabilization and thermoactivation of thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J., Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki ,Y. and Hayashizaki,Y.
Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)
Carninci,P. and Hayashizaki,Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)
Please visit our web site (http://genome.rtc.riken.go.jp) for further details.

FEATURES Location/Qualifiers
source 1. .238
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="B230316K05"
/clone_lib="RIKEN full-length enriched, adult male corpora quadrigemina"
/sex="male"
/tissue_type="corpora quadrigemina"
/dev_stage="adult"
/lab_host="DH10B"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 20.0 and subtraction to Rot = 459.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."

BASE COUNT 71 a 43 c 51 g 73 t
ORIGIN

Query Match 87.0%; Score 17.4; DB 130; Length 238;
Best Local Similarity 94.7%; Pred. No. 89;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 gtcacgtgacacagtcaaa 19
||||||||||||| |||||
Db 72 GTCACGTGACACACTCAAA 54

RESULT 3
AI104756
LOCUS AI104756 295 bp mRNA EST 20-JAN-1999
DEFINITION EST214045 Normalized rat heart, Bento Soares Rattus sp. cDNA clone RHECI60 3' end, mRNA sequence.
ACCESSION AI104756
VERSION AI104756.1 GI:3709041
KEYWORDS EST.
SOURCE Rattus sp.
ORGANISM Rattus sp.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 295)
AUTHORS Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J., Kerlavage,A.R. and Adams,M.D.
TITLE Rat Genome Project: Generation of a Rat EST (REST) Catalog & Rat Gene Index
JOURNAL Unpublished (1998)
COMMENT Contact: Lee, NH
The Institute for Genomic Research
9712, Medical Center Drive, Rockville, MD 20850, USA
Tel: (301)-838-3529
Fax: (301)-838-0208
Email: nhlee@tigr.org
Seq primer: M13-21.

FEATURES Location/Qualifiers
source 1. .295
/organism="Rattus sp."
/db_xref="ATCC (inhost):2025643"
/db_xref="taxon:10118"
/clone="RHECI60"
/clone_lib="Normalized rat heart, Bento Soares"
/note="Organ: heart; Vector: pT7T3Pac; Site_1: EcoRI; Site_2: NotI"

BASE COUNT 87 a 49 c 74 g 85 t
ORIGIN

Query Match 87.0%; Score 17.4; DB 16; Length 295;
Best Local Similarity 94.7%; Pred. No. 92;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 tcacgtgacacagtcaaac 20
||||||||||||| ||||||
Db 233 TCACGTGACACAATCAAAC 251

RESULT 4
AZ015172
LOCUS AZ015172 438 bp DNA GSS 25-FEB-2000
DEFINITION RPCI-23-289M15.TJ RPCI-23 Mus musculus genomic clone RPCI-23-289M15, DNA sequence.
ACCESSION AZ015172
VERSION AZ015172.1 GI:7090556
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 438)
AUTHORS Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret ,B., Levins,M., Mcgann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P. and Fraser,C.M.
TITLE Mouse BAC End Sequences from Library RPCI-23
JOURNAL Unpublished (1999)
COMMENT Other_GSSs: RPCI-23-289M15.TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm) or from Resea ch Genetics (info@resgen.com). BAC end page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 289 row: M column: 15
Seq primer: SP6
Class: BAC ends.

FEATURES Location/Qualifiers
source 1. .438
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-23-289M15"

/clone_lib="RPCI-23"
/sex="Female"
/lab_host="DH10B"
/note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACe3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies). "
BASE COUNT 128 a 90 c 96 g 119 t 5 others
ORIGIN

Query Match 87.0%; Score 17.4; DB 237; Length 438;
Best Local Similarity 94.7%; Pred. No. 97;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 gtcacgtgacacagtcaaa 19
|||| ||||||||||||||
Db 288 GTCATGTGACACAGTCAAA 306

RESULT 5
AQ798101
LOCUS AQ798101 482 bp DNA GSS 09-AUG-1999
DEFINITION HS_3158_A1_H09_MR CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=3158 Col=17 Row=O, DNA sequence.
ACCESSION AQ798101
VERSION AQ798101.1 GI:5715433
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 482)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones may be purchased from Research Genetics (info@resgen.com).
BAC end Web Server: http://www.htsc.washington.edu
Plate: 3158 row: O column: 17
Seq primer: M13 Reverse
Class: BAC ends
High quality sequence stop: 482.
FEATURES Location/Qualifiers
source 1. .482
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=3158 Col=17 Row=O"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in E-Coli DH10B"
BASE COUNT 176 a 78 c 78 g 149 t 1 others
ORIGIN

Query Match 87.0%; Score 17.4; DB 233; Length 482;
Best Local Similarity 94.7%; Pred. No. 98;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 tcacgtgacacagtcaaac 20
||| ||||||||||||||
Db 6 TCATGTGACACAGTCAAAC 24

RESULT 6
AI180191/c
LOCUS AI180191 554 bp mRNA EST 08-JAN-1999
DEFINITION EST223932 Normalized rat spleen, Bento Soares Rattus sp. cDNA clone RSPCR68 3' end, mRNA sequence.
ACCESSION AI180191
VERSION AI180191.1 GI:4135223
KEYWORDS EST.
SOURCE Rattus sp.
ORGANISM Rattus sp.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 554)
AUTHORS Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J., Kerlavage,A.R. and Adams,M.D.
TITLE Rat Genome Project: Generation of a Rat EST (REST) Catalog & Rat Gene Index
JOURNAL Unpublished (1998)
COMMENT On Oct 8, 1998 this sequence version replaced gi:3730829.
Contact: Lee, NH
The Institute for Genomic Research
9712, Medical Center Drive, Rockville, MD 20850, USA
Tel: (301)-838-3529
Fax: (301)-838-0208
Email: nhlee@tigr.org
Seq primer: M13-21.
FEATURES Location/Qualifiers
source 1. .554
/organism="Rattus sp."
/db_xref="taxon:10118"
/clone="RSPCR68"
/clone_lib="Normalized rat spleen, Bento Soares"
/note="Organ: spleen; Vector: pT7T3Pac; Site_1: EcoRI; Site_2: NotI"
BASE COUNT 224 a 101 c 87 g 142 t
ORIGIN

Query Match 87.0%; Score 17.4; DB 17; Length 554;
Best Local Similarity 94.7%; Pred. No. 1e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 tcacgtgacacagtcaaac 20
|||||||||||| ||||||
Db 153 TCACGTGACACAATCAAAC 135

RESULT 7
FR0039113/c
LOCUS FR0039113 258 bp DNA GSS 22-OCT-1999
DEFINITION Fugu rubripes GSS sequence, clone 090H03dE7, genomic survey sequence.
ACCESSION AL126611
VERSION AL126611.1 GI:6108226
KEYWORDS GSS; genome survey sequence.
SOURCE Takifugu rubripes.
ORGANISM Takifugu rubripes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Takifugu.
REFERENCE 1 (bases 1 to 258)
AUTHORS Elgar,G., Clark,M.S., Smith,S., Meek,S., Warner,S., Edwards,Y.J.K., Umrania,Y., Williams,G. and Brenner,S.
TITLE Direct Submission
JOURNAL Submitted (11-OCT-1999) MRC Human Genome Mapping Project Resource Centre, Hinxton, Cambridge, CB10 1SB, UK Email:

```
            biohelp@hgmp.mrc.ac.uk
COMMENT     Vector: pBluescript II KS
            V_type: phagemid
            PRIMER: KS
            DESCR:
            One pass dye-terminator sequencing of cosmid cloned genomic
            sequence.
FEATURES             Location/Qualifiers
     source          1. .258
                     /organism="Takifugu rubripes"
                     /db_xref="taxon:31033"
                     /clone_lib="cosmid 090H03"
                     /clone="090H03dE7"
BASE COUNT       67 a     66 c     58 g     64 t      3 others
ORIGIN


  Query Match             84.0%;  Score 16.8;  DB 222;  Length 258;
  Best Local Similarity   90.0%;  Pred. No. 1.9e+02;
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 gtcacgtgacacagtcaaac 20
          ||||| |||||||||||| |
Db    200 GTCACATGACACAGTCAACC 181


RESULT   8
FR0039183/c
LOCUS       FR0039183     400 bp    DNA             GSS       22-OCT-1999
DEFINITION  Fugu rubripes GSS sequence, clone 090H03fD3, genomic survey
            sequence.
ACCESSION   AL126681
VERSION     AL126681.1  GI:6108296
KEYWORDS    GSS; genome survey sequence.
SOURCE      Takifugu rubripes.
  ORGANISM  Takifugu rubripes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Takifugu.
REFERENCE   1  (bases 1 to 400)
  AUTHORS   Elgar,G., Clark,M.S., Smith,S., Meek,S., Warner,S., Edwards,Y.J.K.,
            Umrania,Y., Williams,G. and Brenner,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-OCT-1999) MRC Human Genome Mapping Project Resource
            Centre, Hinxton, Cambridge, CB10 1SB. UK Email:
            biohelp@hgmp.mrc.ac.uk
COMMENT     Vector: pBluescript II KS
            V_type: phagemid
            PRIMER: KS
            DESCR:
            One pass dye-terminator sequencing of cosmid cloned genomic
            sequence.
FEATURES             Location/Qualifiers
     source          1. .400
                     /organism="Takifugu rubripes"
                     /db_xref="taxon:31033"
                     /clone_lib="cosmid 090H03"
                     /clone="090H03fD3"
BASE COUNT      102 a    102 c     88 g    103 t      5 others
ORIGIN


  Query Match             84.0%;  Score 16.8;  DB 222;  Length 400;
  Best Local Similarity   90.0%;  Pred. No. 2e+02;
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 gtcacgtgacacagtcaaac 20
          ||||| |||||||||||| |
Db    296 GTCACATGACACAGTCAACC 277


RESULT   9
FR0039084/c
LOCUS       FR0039084     513 bp    DNA             GSS       22-OCT-1999
DEFINITION  Fugu rubripes GSS sequence, clone 090H03cA5, genomic survey
            sequence.
ACCESSION   AL126582
VERSION     AL126582.1  GI:6108197
KEYWORDS    GSS; genome survey sequence.
SOURCE      Takifugu rubripes.
  ORGANISM  Takifugu rubripes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Takifugu.
REFERENCE   1  (bases 1 to 513)
  AUTHORS   Elgar,G., Clark,M.S., Smith,S., Meek,S., Warner,S., Edwards,Y.J.K.,
            Umrania,Y., Williams,G. and Brenner,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-OCT-1999) MRC Human Genome Mapping Project Resource
            Centre, Hinxton, Cambridge, CB10 1SB. UK Email:
            biohelp@hgmp.mrc.ac.uk
COMMENT     Vector: pBluescript II KS
            V_type: phagemid
            PRIMER: KS
            DESCR:
            One pass dye-terminator sequencing of cosmid cloned genomic
            sequence.
FEATURES             Location/Qualifiers
     source          1. .513
                     /organism="Takifugu rubripes"
                     /db_xref="taxon:31033"
                     /clone_lib="cosmid 090H03"
                     /clone="090H03cA5"
BASE COUNT      132 a    130 c    112 g    134 t      5 others
ORIGIN


  Query Match             84.0%;  Score 16.8;  DB 222;  Length 513;
  Best Local Similarity   90.0%;  Pred. No. 2e+02;
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 gtcacgtgacacagtcaaac 20
          ||||| |||||||||||| |
Db     30 GTCACATGACACAGTCAACC 11


RESULT  10
BE368176
LOCUS       BE368176      539 bp    mRNA            EST       21-JUL-2000
DEFINITION  601222089F1 NCI_CGAP_Lu29 Mus musculus cDNA clone IMAGE:3590711 5',
            mRNA sequence.
ACCESSION   BE368176
VERSION     BE368176.1  GI:9313539
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 539)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM8759  row: b  column: 24
            High quality sequence stop: 536.
FEATURES             Location/Qualifiers
```

```
     source          1. .539
                     /organism="Mus musculus"
                     /strain="CZECH II (feral)"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:3590711"
                     /clone_lib="NCI_CGAP_Lu29"
                     /tissue_type="spontaneous tumor, metastatic to mammary.
                     Stem cell origin."
                     /lab_host="DH10B"
                     /note="Organ: lung; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                     Library constructed by Life Technologies. Investigator
                     providing samples: Gilbert Smith, NIH"
BASE COUNT      201 a     92 c    143 g    103 t
ORIGIN


  Query Match             84.0%;  Score 16.8;  DB 166;  Length 539;
  Best Local Similarity   90.0%;  Pred. No. 2.1e+02;
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 gtcacgtgacacagtcaaac 20
          |||||||||||||||  || |
Db    207 GTCACGTGACACAGTAAAGC 226


RESULT  11
BF500440/c
LOCUS       BF500440      548 bp    mRNA            EST       16-APR-2001
DEFINITION  AT15369.5prime AT Drosophila melanogaster adult testes pOTB7
            Drosophila melanogaster cDNA clone AT15369 5 similar to CG9778:
            FBan0009778 last_updated:000321, mRNA sequence.
ACCESSION   BF500440
VERSION     BF500440.1  GI:11583741
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 548)
  AUTHORS   Stapleton,M., Brokstein,P., Hong,L., Agbayani,A., Baxter,E., Berman
            ,B., Carlson,J., Champe,M., Chavez,C., Chew,M., Dorsett,V., Farfan
            ,D., Frise,E., George,R., Gonzalez,M., Guarin,H., Harris,N., Li,P.,
            Liao,G., Miranda,A., Misra,S., Mungall,C.J., Nunoo,J., Pacleb,J.,
            Park,S., Paragas,V., Phouanenavong,S., Wan,K., Yu,C., Lewis,S.E.,
            Celniker,S. and Rubin,G.M.
  TITLE     Berkeley Drosophila Gene Collection Project
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Stapleton, M.
            BDGP
            Lawrence Berkeley National Lab
            One Cyclotron Rd, Berkeley, CA 94720, USA
            Fax: 510 486 6798
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            hit genomic sequence AE003606
            Plate: AT.153  row: F  column: 9
            High quality sequence stop: 513.
FEATURES             Location/Qualifiers
     source          1. .548
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="AT15369"
                     /clone_lib="AT Drosophila melanogaster adult testes pOTB7"
                     /sex="male"
                     /dev_stage="0-3 day old Ore-R males"
                     /lab_host="Plates AT.10-AT.120: DH5-alpha. Plates
                     AT.121-AT.319: DH5-alpha TonA"
                     /note="Organ: ADULT testes; Vector: pOTB7; Site_1: EcoRI;
                     Site_2: XhoI; The mRNA for the testis library was made
                     from testes and seminal vessicles hand dissected from 0-3
                     day old Ore-R males. RNA kindly provided by the lab of
                     Margaret Fuller. Sized fractionated cDNAs were directly
                     ligated into pOTB7. Plasmid cDNA library."
BASE COUNT      115 a     155 c     152 g     126 t
ORIGIN


  Query Match             84.0%;  Score 16.8;  DB 149;  Length 548;
  Best Local Similarity   90.0%;  Pred. No. 2.1e+02;
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 gtcacgtgacacagtcaaac 20
          |||||||||||||  |||| ||
Db    537 GTCACGTGACACCGTCACAC 518


RESULT  12
W73212/c
LOCUS       W73212        562 bp    mRNA            EST       16-OCT-1996
DEFINITION  zd53e10.r1 Soares_fetal_heart_NbHH19W Homo sapiens cDNA clone
            IMAGE:344394 5', mRNA sequence.
ACCESSION   W73212
VERSION     W73212.1  GI:1383346
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 562)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman
            ,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J.,
            Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston
            ,R., Williamson,A., Wohldmann,P. and Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 623   Std Error: 0.00
            Seq primer: ETPrimer
            High quality sequence stop: 437.
FEATURES             Location/Qualifiers
     source          1. .562
                     /organism="Homo sapiens"
                     /db_xref="GDB:1269769"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:344394"
                     /clone_lib="Soares_fetal_heart_NbHH19W"
                     /sex="unknown"
                     /dev_stage="19 weeks"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Organ: heart; Vector: pT7T3D (Pharmacia) with a
                     modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCATCTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was size selected, ligated to Eco RI
                     adapters (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of a modified pT7T3 vector
                     (Pharmacia). Library went through one round of
                     normalization to a Cot = 5. Library constructed by
                     M.Fatima Bonaldo. This library was constructed from the
                     same fetus as the fetal lung library, Soares fetal lung
                     NbHL19W."
BASE COUNT      121 a     150 c     147 g     143 t       1 others
ORIGIN


  Query Match             84.0%;  Score 16.8;  DB 190;  Length 562;
  Best Local Similarity   90.0%;  Pred. No. 2.1e+02;
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;
```

```
Qy      1 gtcacgtgacacagtcaaac 20
          |||||| ||||||| |||||
Db    549 GTCACGAGACACAGCCAAAC 530


RESULT 13
BF232638/c
LOCUS       BF232638       804 bp    mRNA            EST       14-NOV-2000
DEFINITION  602023221F1 NCI_CGAP_Li9 Mus musculus cDNA clone IMAGE:4158796 5',
            mRNA sequence.
ACCESSION   BF232638
VERSION     BF232638.1  GI:11142204
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 804)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9435  row: i  column: 05
            High quality sequence stop: 657.
FEATURES             Location/Qualifiers
     source          1. .804
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4158796"
                     /clone_lib="NCI_CGAP_Li9"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: liver; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 1.9 kb. Constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT      196 a    216 c    233 g    159 t
ORIGIN


  Query Match            84.0%;  Score 16.8;  DB 146;  Length 804;
  Best Local Similarity  90.0%;  Pred. No. 2.2e+02;
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 gtcacgtgacacagtcaaac 20
          || |||||||||| ||||||
Db    455 GTGACGTGACACATTCAAAC 436


RESULT 14
CNS02G0P/c
LOCUS       CNS02G0P       835 bp    DNA             GSS       13-MAY-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence T7 end of clone
            135P15 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL195730
VERSION     AL195730.1  GI:7833880
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 835)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
  TITLE     Charaterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 835)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 835)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .835
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="135P15"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0AG135CH08LP1~end : T7"
BASE COUNT      192 a    195 c    246 g    200 t      2 others
ORIGIN


  Query Match            84.0%;  Score 16.8;  DB 220;  Length 835;
  Best Local Similarity  90.0%;  Pred. No. 2.2e+02;
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 gtcacgtgacacagtcaaac 20
          ||||| ||||| ||||||||
Db    531 GTCACATGACAGAGTCAAAC 512


RESULT 15
CNS03EDV
LOCUS       CNS03EDV       835 bp    DNA             GSS       17-MAY-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone
            020G19 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL240268
VERSION     AL240268.1  GI:7961037
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 835)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
  TITLE     Charaterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 835)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 835)
  AUTHORS   Genoscope.
```

```
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .835
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="020G19"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0BG020AD10SP1-end :
                     PUC-Ori"
BASE COUNT      185 a    245 c    233 g    168 t      4 others
ORIGIN


  Query Match             84.0%;  Score 16.8;  DB 221;  Length 835;
  Best Local Similarity   90.0%;  Pred. No. 2.2e+02;
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 gtcacgtgacacagtcaaac 20
           |||||||||||: ||| ||||
Db     261 GTCACGTGACAAAGTGAAAC 280


Search completed: July 25, 2001, 04:54:54
Job time: 10427 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: July 25, 2001, 05:17:09 ; Search time 1290.33 Seconds
(without alignments)
227.761 Million cell updates/sec

Title: US-09-142-095-4
Perfect score: 19
Sequence: 1 tttgctcctgccagaggtt 19

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues

Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_ba3:*
4: gb_in1:*
5: gb_in2:*
6: gb_in3:*
7: gb_om:*
8: gb_ov:*
9: gb_pat1:*
10: gb_pat2:*
11: gb_ph:*
12: gb_pl1:*
13: gb_pl2:*
14: gb_pl3:*
15: gb_pl4:*
16: em_ba1:*
17: em_ba2:*
18: em_fun:*
19: em_htgo_hum:*
20: em_htgo_inv:*
21: em_htgo_rod:*
22: em_htg_hum1:*
23: em_htg_hum2:*
24: em_htg_hum3:*
25: em_htg_hum4:*
26: em_htg_hum5:*
27: em_htg_hum6:*
28: em_htg_hum7:*
29: em_htg_hum8:*
30: em_htg_inv1:*
31: em_htg_inv2:*
32: em_htg_other:*
33: em_htg_rod:*
34: em_hum1:*
35: em_hum2:*
36: em_hum3:*
37: em_hum4:*
38: em_hum5:*
39: em_hum6:*
40: em_hum7:*
41: em_in:*
42: em_om:*
43: em_or:*
44: em_ov:*
45: em_pat:*
46: em_ph:*
47: em_pl:*
48: em_ro:*
49: em_sts:*
50: em_sy:*
51: em_un:*
52: em_vi:*
53: gb_sts1:*
54: gb_sts2:*
55: gb_sts3:*
56: gb_sy:*
57: gb_un:*
58: gb_vi1:*
59: gb_vi2:*
60: gb_htg1:*
61: gb_htg2:*
62: gb_htg3:*
63: gb_htg4:*
64: gb_htg5:*
65: gb_htg6:*
66: gb_htg7:*
67: gb_htg8:*
68: gb_htg9:*
69: gb_htg10:*
70: gb_htg11:*
71: gb_htg12:*
72: gb_htg13:*
73: gb_htg14:*
74: gb_htg15:*
75: gb_htg16:*
76: gb_htg17:*
77: gb_htg18:*
78: gb_htg19:*
79: gb_htg20:*
80: gb_htg21:*
81: gb_htg22:*
82: gb_htg23:*
83: gb_htg24:*
84: gb_htg25:*
85: gb_pr1:*
86: gb_pr2:*
87: gb_pr3:*
88: gb_pr4:*
89: gb_pr5:*
90: gb_pr6:*
91: gb_pr7:*
92: gb_pr8:*
93: gb_pr9:*
94: gb_ro1:*
95: gb_ro2:*
96: gb_in4:*
97: gb_pr10:*
98: em_ba3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		19	100.0	19	9	A65503	A65503 Sequence 4
c 2		19	100.0	177	89	AF135465	AF135465 Gorilla g
c 3		19	100.0	200	89	AF135466	AF135466 Pongo pyg
c 4		19	100.0	207	89	AF135470	AF135470 Cebus ape
c 5		19	100.0	208	89	AF135464	AF135464 Gorilla g
c 6		19	100.0	208	89	AF135471	AF135471 Saimiri b
c 7		19	100.0	234	89	AF135462	AF135462 Pan panis
c 8		19	100.0	234	89	AF135463	AF135463 Pan trogl

c	9	19	100.0	292	89	AF357220	AF357220 Homo sapi
c	10	19	100.0	531	89	AF352795	AF352795 Homo sapi
c	11	19	100.0	541	89	AF180372	AF180372 Homo sapi
c	12	19	100.0	620	9	A65504	A65504 Sequence 5
c	13	19	100.0	918	88	AF110194	AF110194 Homo sapi
c	14	19	100.0	1190	97	HUMUGT1A	M84125 Human bilir
c	15	19	100.0	3341	91	D87674	D87674 Homo sapien
	16	19	100.0	68770	86	AC006985	AC006985 Homo sapi
	17	19	100.0	176619	70	AC026497	AC026497 Homo sapi
c	18	19	100.0	198872	89	AF297093	AF297093 Homo sapi
c	19	17.4	91.6	208	89	AF135467	AF135467 Trachypit
c	20	17.4	91.6	208	89	AF135468	AF135468 Papio cyn
	21	17.4	91.6	141347	76	AC083820	AC083820 Rattus no
	22	17.4	91.6	158091	87	AC018904	AC018904 Homo sapi
c	23	17.4	91.6	165730	72	AC061988	AC061988 Homo sapi
c	24	17	89.5	159762	62	AC012040	AC012040 Homo sapi
c	25	17	89.5	190184	88	AC026464	AC026464 Homo sapi
c	26	17	89.5	196518	70	AC026474	AC026474 Homo sapi
	27	17	89.5	208378	76	AC079988	AC079988 Homo sapi
	28	16.4	86.3	910	1	AF134384	AF134384 Unculture
c	29	16.4	86.3	920	2	AF293014	AF293014 Unculture
c	30	16.4	86.3	920	2	AF293015	AF293015 Unculture
	31	16.4	86.3	1338	3	UEAJ9508	AJ009508 unculture
	32	16.4	86.3	1340	3	UEAJ9509	AJ009509 unculture
	33	16.4	86.3	1350	3	MSU89773	U89773 Methnosarci
	34	16.4	86.3	1352	2	AF229776	AF229776 Unculture
	35	16.4	86.3	1362	3	MVU20150	U20150 Methanosarc
	36	16.4	86.3	1420	3	MTU20154	U20154 Methanolobu
	37	16.4	86.3	1421	3	MMU20151	U20151 Methanosarc
	38	16.4	86.3	1422	3	MTHRR16S	M59133 Methanohalo
	39	16.4	86.3	1423	3	MTARR16S	M59141 Methanosaet
	40	16.4	86.3	1425	3	MESRR16SB	M59138 Methanosarc
	41	16.4	86.3	1426	3	MESRR16S	M59136 Methanosarc
	42	16.4	86.3	1426	3	MEURR16S	M59132 Methanophil
	43	16.4	86.3	1426	3	MSU20153	U20153 Methanosarc
	44	16.4	86.3	1427	3	MESRR16SA	M59137 Methanosarc
	45	16.4	86.3	1428	3	MESRR16SG	M59140 Methanosarc

ALIGNMENTS

```
RESULT   1
A65503
LOCUS       A65503          19 bp    DNA             PAT       29-MAR-1999
DEFINITION  Sequence 4 from Patent WO9732042.
ACCESSION   A65503
VERSION     A65503.1  GI:4531238
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 19)
  AUTHORS   Burchell,B.
  TITLE     DRUG TRIAL ASSAY SYSTEM
  JOURNAL   Patent: WO 9732042-A 4 04-SEP-1997;
            UNIV DUNDEE (GB)
COMMENT     Other publication AU 2224197 19970916.
FEATURES             Location/Qualifiers
     source          1. .19
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT        2 a      5 c      5 g      7 t
ORIGIN


  Query Match           100.0%;  Score 19;  DB 9;  Length 19;
  Best Local Similarity  100.0%;  Pred. No. 17;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tttgctcctgccagaggtt 19
          |||||||||||||||||||
Db      1 TTTGCTCCTGCCAGAGGTT 19
```

```
RESULT   2
AF135465/c
LOCUS       AF135465       177 bp    DNA             PRI       21-NOV-1999
DEFINITION  Gorilla gorilla UDP-glucuronosyltransferase 1A1 (UGT1A1) gene,
            promoter region and partial cds.
ACCESSION   AF135465
VERSION     AF135465.1  GI:6456547
KEYWORDS    .
SOURCE      gorilla.
  ORGANISM  Gorilla gorilla
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Gorilla.
REFERENCE   1  (bases 1 to 177)
  AUTHORS   Hall,D., Ybazeta,G., Destro-Bisol,G., Petzl-Erler,M.L. and Di
            Rienzo,A.
  TITLE     Variability at the uridine diphosphate glucuronosyltransferase 1A1
            promoter in human populations and primates
  JOURNAL   Pharmacogenetics (1999) In press
REFERENCE   2  (bases 1 to 177)
  AUTHORS   Ybazeta,G., Hall,D. and Di Rienzo,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-MAR-1999) Human Genetics, University of Chicago, 924
            E. 57th Street, Chicago, IL 60637, USA
FEATURES             Location/Qualifiers
     source          1. .177
                     /organism="Gorilla gorilla"
                     /db_xref="taxon:9593"
                     /note="from the Brookfield Zoo"
     mRNA            <91. .>177
                     /gene="UGT1A1"
                     /product="UDP-glucuronosyltransferase 1A1"
     gene            <91. .>177
                     /gene="UGT1A1"
     CDS             91. .>177
                     /gene="UGT1A1"
                     /codon_start=1
                     /product="UDP-glucuronosyltransferase 1A1"
                     /protein_id="AAF09176.1"
                     /db_xref="GI:6456548"
                     /translation="MAVESQGGHPLVLGLLLCVLGPVVSHAGK"
BASE COUNT       37 a     39 c     57 g     44 t
ORIGIN


  Query Match           100.0%;  Score 19;  DB 89;  Length 177;
  Best Local Similarity  100.0%;  Pred. No. 15;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tttgctcctgccagaggtt 19
          |||||||||||||||||||
Db     84 TTTGCTCCTGCCAGAGGTT 66
```

```
RESULT   3
AF135466/c
LOCUS       AF135466       200 bp    DNA             PRI       21-NOV-1999
DEFINITION  Pongo pygmaeus UDP-glucuronosyltransferase 1A1 (UGT1A1) gene,
            promoter region and partial cds.
ACCESSION   AF135466
VERSION     AF135466.1  GI:6456549
KEYWORDS    .
SOURCE      orangutan.
  ORGANISM  Pongo pygmaeus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pongo.
REFERENCE   1  (bases 1 to 200)
  AUTHORS   Hall,D., Ybazeta,G., Destro-Bisol,G., Petzl-Erler,M.L. and Di
            Rienzo,A.
  TITLE     Variability at the uridine diphosphate glucuronosyltransferase 1A1
            promoter in human populations and primates
  JOURNAL   Pharmacogenetics (1999) In press
```

```
REFERENCE   2  (bases 1 to 200)
  AUTHORS   Ybazeta,G., Hall,D. and Di Rienzo,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-MAR-1999) Human Genetics, University of Chicago, 924
            E. 57th Street, Chicago, IL 60637, USA
FEATURES             Location/Qualifiers
     source          1. .200
                     /organism="Pongo pygmaeus"
                     /db_xref="taxon:9600"
     mRNA            <121. .>200
                     /gene="UGT1A1"
                     /product="UDP-glucuronosyltransferase 1A1"
     gene            <121. .>200
                     /gene="UGT1A1"
     CDS             121. .>200
                     /gene="UGT1A1"
                     /codon_start=1
                     /product="UDP-glucuronosyltransferase 1A1"
                     /protein_id="AAF09177.1"
                     /db_xref="GI:6456550"
                     /translation="MAVESQGGRPLVLGLLLCVLGPVVSHA"
BASE COUNT       40 a     46 c     62 g     52 t
ORIGIN


  Query Match           100.0%;  Score 19;  DB 89;  Length 200;
  Best Local Similarity  100.0%;  Pred. No. 15;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tttgctcctgccagaggtt 19
          |||||||||||||||||||
Db    114 TTTGCTCCTGCCAGAGGTT 96


RESULT  4
AF135470/c
LOCUS       AF135470      207 bp    DNA             PRI       21-NOV-1999
DEFINITION  Cebus apella UDP-glucuronosyltransferase 1A1 (UGT1A1) gene,
            promoter region and partial cds.
ACCESSION   AF135470
VERSION     AF135470.1  GI:6456557
KEYWORDS    .
SOURCE      brown capuchin.
  ORGANISM  Cebus apella
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Platyrrhini; Cebidae; Cebinae; Cebus.
REFERENCE   1  (bases 1 to 207)
  AUTHORS   Hall,D., Ybazeta,G., Destro-Bisol,G., Petzl-Erler,M.L. and Di
            Rienzo,A.
  TITLE     Variability at the uridine diphosphate glucuronosyltransferase 1A1
            promoter in human populations and primates
  JOURNAL   Pharmacogenetics (1999) In press
REFERENCE   2  (bases 1 to 207)
  AUTHORS   Ybazeta,G., Hall,D. and Di Rienzo,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-MAR-1999) Human Genetics, University of Chicago, 924
            E. 57th Street, Chicago, IL 60637, USA
FEATURES             Location/Qualifiers
     source          1. .207
                     /organism="Cebus apella"
                     /db_xref="taxon:9515"
     mRNA            <144. .>207
                     /gene="UGT1A1"
                     /product="UDP-glucuronosyltransferase 1A1"
     gene            <144. .>207
                     /gene="UGT1A1"
     CDS             144. .>207
                     /gene="UGT1A1"
                     /codon_start=1
                     /product="UDP-glucuronosyltransferase 1A1"
                     /protein_id="AAF09181.1"
                     /db_xref="GI:6456558"
                     /translation="MPACPGPAAVCAGPGSVPCWE"
BASE COUNT       43 a     46 c     62 g     56 t
ORIGIN


  Query Match           100.0%;  Score 19;  DB 89;  Length 207;
  Best Local Similarity  100.0%;  Pred. No. 15;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tttgctcctgccagaggtt 19
          |||||||||||||||||||
Db    114 TTTGCTCCTGCCAGAGGTT 96


RESULT  5
AF135464/c
LOCUS       AF135464      208 bp    DNA             PRI       21-NOV-1999
DEFINITION  Gorilla gorilla UDP-glucuronosyltransferase 1A1 (UGT1A1) gene,
            promoter region and partial cds.
ACCESSION   AF135464
VERSION     AF135464.1  GI:6456545
KEYWORDS    .
SOURCE      gorilla.
  ORGANISM  Gorilla gorilla
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Gorilla.
REFERENCE   1  (bases 1 to 208)
  AUTHORS   Hall,D., Ybazeta,G., Destro-Bisol,G., Petzl-Erler,M.L. and Di
            Rienzo,A.
  TITLE     Variability at the uridine diphosphate glucuronosyltransferase 1A1
            promoter in human populations and primates
  JOURNAL   Pharmacogenetics (1999) In press
REFERENCE   2  (bases 1 to 208)
  AUTHORS   Ybazeta,G., Hall,D. and Di Rienzo,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-MAR-1999) Human Genetics, University of Chicago, 924
            E. 57th Street, Chicago, IL 60637, USA
FEATURES             Location/Qualifiers
     source          1. .208
                     /organism="Gorilla gorilla"
                     /db_xref="taxon:9593"
                     /note="from Yerkes Regional Primate Center"
     mRNA            <133. .>208
                     /gene="UGT1A1"
                     /product="UDP-glucuronosyltransferase 1A1"
     gene            <133. .>208
                     /gene="UGT1A1"
     CDS             133. .>208
                     /gene="UGT1A1"
                     /codon_start=1
                     /product="UDP-glucuronosyltransferase 1A1"
                     /protein_id="AAF09175.1"
                     /db_xref="GI:6456546"
                     /translation="MAVESQGGHPLVLGLLLCVLGPVVS"
BASE COUNT       43 a     47 c     62 g     56 t
ORIGIN


  Query Match           100.0%;  Score 19;  DB 89;  Length 208;
  Best Local Similarity  100.0%;  Pred. No. 15;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tttgctcctgccagaggtt 19
          |||||||||||||||||||
Db    126 TTTGCTCCTGCCAGAGGTT 108


RESULT  6
AF135471/c
LOCUS       AF135471      208 bp    DNA             PRI       21-NOV-1999
DEFINITION  Saimiri boliviensis UDP-glucuronosyltransferase 1A1 (UGT1A1) gene,
            promoter region and partial cds.
ACCESSION   AF135471
VERSION     AF135471.1  GI:6456559
```

```
KEYWORDS    .
SOURCE      Bolivian squirrel monkey.
  ORGANISM  Saimiri boliviensis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Platyrrhini; Cebidae; Cebinae;
            Saimiri.
REFERENCE   1  (bases 1 to 208)
  AUTHORS   Hall,D., Ybazeta,G., Destro-Bisol,G., Petzl-Erler,M.L. and Di
            Rienzo,A.
  TITLE     Variability at the uridine diphosphate glucuronosyltransferase 1A1
            promoter in human populations and primates
  JOURNAL   Pharmacogenetics (1999) In press
REFERENCE   2  (bases 1 to 208)
  AUTHORS   Ybazeta,G., Hall,D. and Di Rienzo,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-MAR-1999) Human Genetics, University of Chicago, 924
            E. 57th Street, Chicago, IL 60637, USA
FEATURES             Location/Qualifiers
     source          1. .208
                     /organism="Saimiri boliviensis"
                     /db_xref="taxon:27679"
     mRNA            <142. .>208
                     /gene="UGT1A1"
                     /product="UDP-glucuronosyltransferase 1A1"
     gene            <142. .>208
                     /gene="UGT1A1"
     CDS             142. .>208
                     /gene="UGT1A1"
                     /codon_start=1
                     /product="UDP-glucuronosyltransferase 1A1"
                     /protein_id="AAF09182.1"
                     /db_xref="GI:6456560"
                     /translation="MAAESRGGCQLVLGLLLCVLGP"
BASE COUNT       44 a     52 c     58 g     54 t
ORIGIN


  Query Match           100.0%;  Score 19;  DB 89;  Length 208;
  Best Local Similarity  100.0%;  Pred. No. 15;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tttgctcctgccagaggtt 19
          |||||||||||||||||||
Db    135 TTTGCTCCTGCCAGAGGTT 117


RESULT   7
AF135462/c
LOCUS       AF135462      234 bp    DNA             PRI       21-NOV-1999
DEFINITION  Pan paniscus UDP-glucuronosyltransferase 1A1 (UGT1A1) gene,
            promoter region and partial cds.
ACCESSION   AF135462
VERSION     AF135462.1  GI:6456541
KEYWORDS    .
SOURCE      pygmy chimpanzee.
  ORGANISM  Pan paniscus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1  (bases 1 to 234)
  AUTHORS   Hall,D., Ybazeta,G., Destro-Bisol,G., Petzl-Erler,M.L. and Di
            Rienzo,A.
  TITLE     Variability at the uridine diphosphate glucuronosyltransferase 1A1
            promoter in human populations and primates
  JOURNAL   Pharmacogenetics (1999) In press
REFERENCE   2  (bases 1 to 234)
  AUTHORS   Ybazeta,G., Hall,D. and Di Rienzo,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-MAR-1999) Human Genetics, University of Chicago, 924
            E. 57th Street, Chicago, IL 60637, USA
FEATURES             Location/Qualifiers
     source          1. .234
                     /organism="Pan paniscus"
                     /db_xref="taxon:9597"
     mRNA            <148. .>234
                     /gene="UGT1A1"
                     /product="UDP-glucuronosyltransferase 1A1"
     gene            <148. .>234
                     /gene="UGT1A1"
     CDS             148. .>234
                     /gene="UGT1A1"
                     /codon_start=1
                     /product="UDP-glucuronosyltransferase 1A1"
                     /protein_id="AAF09173.1"
                     /db_xref="GI:6456542"
                     /translation="MAVESQGGRPLVLGLLLCVLGPVVSHAGK"
BASE COUNT       50 a     53 c     69 g     62 t
ORIGIN


  Query Match           100.0%;  Score 19;  DB 89;  Length 234;
  Best Local Similarity  100.0%;  Pred. No. 15;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tttgctcctgccagaggtt 19
          |||||||||||||||||||
Db    141 TTTGCTCCTGCCAGAGGTT 123


RESULT   8
AF135463/c
LOCUS       AF135463      234 bp    DNA             PRI       21-NOV-1999
DEFINITION  Pan troglodytes UDP-glucuronosyltransferase 1A1 (UGT1A1) gene,
            promoter region and partial cds.
ACCESSION   AF135463
VERSION     AF135463.1  GI:6456543
KEYWORDS    .
SOURCE      chimpanzee.
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1  (bases 1 to 234)
  AUTHORS   Hall,D., Ybazeta,G., Destro-Bisol,G., Petzl-Erler,M.L. and Di
            Rienzo,A.
  TITLE     Variability at the uridine diphosphate glucuronosyltransferase 1A1
            promoter in human populations and primates
  JOURNAL   Pharmacogenetics (1999) In press
REFERENCE   2  (bases 1 to 234)
  AUTHORS   Ybazeta,G., Hall,D. and Di Rienzo,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-MAR-1999) Human Genetics, University of Chicago, 924
            E. 57th Street, Chicago, IL 60637, USA
FEATURES             Location/Qualifiers
     source          1. .234
                     /organism="Pan troglodytes"
                     /db_xref="taxon:9598"
     mRNA            <148. .>234
                     /gene="UGT1A1"
                     /product="UDP-glucuronosyltransferase 1A1"
     gene            <148. .>234
                     /gene="UGT1A1"
     CDS             148. .>234
                     /gene="UGT1A1"
                     /codon_start=1
                     /product="UDP-glucuronosyltransferase 1A1"
                     /protein_id="AAF09174.1"
                     /db_xref="GI:6456544"
                     /translation="MAVESQGGRPLVLGLLLCVLGPVVSHAGK"
BASE COUNT       50 a     53 c     69 g     62 t
ORIGIN


  Query Match           100.0%;  Score 19;  DB 89;  Length 234;
  Best Local Similarity  100.0%;  Pred. No. 15;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tttgctcctgccagaggtt 19
```

```
          ||||||||||||||||||
Db    141 TTTGCTCCTGCCAGAGGTT 123


RESULT  9
AF357220/c
LOCUS       AF357220      292 bp    DNA             PRI       27-MAR-2001
DEFINITION  Homo sapiens bilirubin UDP-glucuronosyltransferase 1 (UGT1) gene,
            UGT1*1 allele, promoter and partial cds.
ACCESSION   AF357220
VERSION     AF357220.1  GI:13448828
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 292)
  AUTHORS   McKie,K., Addington,T., Nguyen,T.S., Glendenning,M., Kutlar,F. and
            Kutlar,A.
  TITLE     Detection of TATA box TA repeat region [6(TA)repeat] of human
            bilirubin UDP-glucuronosyltransferase 1-1 gene (UGT1*1) in an
            African American individual
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 292)
  AUTHORS   McKie,K., Addington,T., Nguyen,T.S., Glendenning,M., Kutlar,F. and
            Kutlar,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-MAR-2001) Medicine/Hemoglobin DNA Laboratory, Medical
            College of Georgia, 15th Street, Augusta, GA 30912, USA
FEATURES             Location/Qualifiers
     source          1. .292
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /map="2q37"
                     /cell_type="WBC"
                     /tissue_type="whole blood"
     gene            1. .>292
                     /gene="UGT1"
                     /allele="UGT1*1"
     promoter        1. .64
                     /gene="UGT1"
     variation       28. .39
                     /gene="UGT1"
                     /note="most common type of repeat; 6(TA)"
     TATA_signal     28. .42
                     /gene="UGT1"
     repeat_region   28. .39
                     /note="polymorphic region"
                     /rpt_type=tandem
                     /rpt_unit=ta
     mRNA            65. .>292
                     /gene="UGT1"
                     /product="bilirubin UDP-glucuronosyltransferase 1"
     CDS             81. .>292
                     /gene="UGT1"
                     /codon_start=1
                     /product="bilirubin UDP-glucuronosyltransferase 1"
                     /protein_id="AAK27223.1"
                     /db_xref="GI:13448829"
                     /translation="MAVESQGGRPLVLGLLLCVLGPVVSHAGKILLIPVDGSHWLSML
                     GAIQQLQQRGHEIVVLAPDASLYIRDG"
BASE COUNT       60 a     69 c     95 g     68 t
ORIGIN


  Query Match           100.0%;  Score 19;  DB 89;  Length 292;
  Best Local Similarity  100.0%;  Pred. No. 14;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tttgctcctgccagaggtt 19
          ||||||||||||||||||||
Db     74 TTTGCTCCTGCCAGAGGTT 56


RESULT  10
AF352795/c
LOCUS       AF352795      531 bp    DNA             PRI       10-APR-2001
DEFINITION  Homo sapiens bilirubin UDP-glucuronosyltrasferase 1-1 (UGT1A1) gene,
            UGT1A1*1 allele, partial cds.
ACCESSION   AF352795
VERSION     AF352795.1  GI:13569708
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 531)
  AUTHORS   McKie,K., Kutlar,F., Glendenning,M. and Kutlar,A.
  TITLE     7(TA) repeat polymorphism of the TATA box of human bilirubin
            UDP-glucuronosyltransferase 1-1(UGT1A1*1) gene in a patient with
            sickle cell anemia + high bilirubinemia
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 531)
  AUTHORS   McKie,K., Kutlar,F., Glendenning,M. and Kutlar,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-FEB-2001) Medicine/Hemoglobin DNA Laboratory, Medical
            College of Georgia, 15th St., AC-1000, Augusta, GA 30912, USA
FEATURES             Location/Qualifiers
     source          1. .531
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /map="2q37"
                     /cell_type="white blood cell"
                     /tissue_type="blood"
     gene            <1. .>531
                     /gene="UGT1A1"
                     /note="UGT1"
                     /allele="UGT1A1*1"
     promoter        <1. .313
                     /gene="UGT1A1"
     TATA_signal     279. .295
                     /gene="UGT1A1"
     repeat_region   279. .>292
                     /note="repeat polymorphism compared to UGT1A1 sequence
                     presented in GenBank Accession Number AF180372; contains 7
                     ta repeats frequently detected in patients with high
                     bilirubinemia"
                     /rpt_type=tandem
                     /rpt_unit=ta
     mRNA            314. .>531
                     /gene="UGT1A1"
                     /product="bilirubin UDP-glucronosyltrasferase 1-1"
     CDS             334. .>531
                     /gene="UGT1A1"
                     /codon_start=1
                     /product="bilirubin UDP-glucronosyltrasferase 1-1"
                     /protein_id="AAK31204.1"
                     /db_xref="GI:13569709"
                     /translation="MAVESQGGRPLVLGLLLCVLGPVVSHAGKILLIPVDGSHWLSML
                     GAIQQLQQRGHEIVVLAPDASL"
BASE COUNT      120 a    121 c    137 g    153 t
ORIGIN


  Query Match           100.0%;  Score 19;  DB 89;  Length 531;
  Best Local Similarity  100.0%;  Pred. No. 14;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tttgctcctgccagaggtt 19
          ||||||||||||||||||||
Db    327 TTTGCTCCTGCCAGAGGTT 309


RESULT  11
```

```
AF180372/c
LOCUS       AF180372      541 bp    DNA             PRI       05-OCT-1999
DEFINITION  Homo sapiens bilirubin UDP-glucuronosyltransferase 1-1 (UGT1) gene,
            UGT1*1 allele, partial cds.
ACCESSION   AF180372
VERSION     AF180372.1  GI:6010649
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 541)
  AUTHORS   Kutlar,F., Sromek,E., Leithner,C., Nechtman,J. and Kutlar,A.
  TITLE     Detection of the TATA box polymorphism of the human bilirubin
            UDP-glucuronosyltransferase 1-1 gene (UGT1*1) in a patient with
            sickle cell anemia
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 541)
  AUTHORS   Kutlar,F., Sromek,E., Leithner,C., Nechtman,J. and Kutlar,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-AUG-1999) Medicine, Hematology/Oncology-Sickle Cell
            Center, Medical College of Georgia, 15th Street, AC-1000, Augusta,
            GA 30912, USA
FEATURES             Location/Qualifiers
     source          1. .541
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /map="2q37"
                     /cell_type="white blood cells"
                     /tissue_type="whole blood"
     gene            1. .>541
                     /gene="UGT1"
                     /note="GNT1; UGT1A1"
                     /allele="UGT1*1"
     promoter        1. .313
                     /gene="UGT1"
     variation       226
                     /gene="UGT1"
                     /replace="t"
     repeat_region   279. .288
                     /note="polymorphic region"
                     /rpt_type=tandem
                     /rpt_unit=ta
     TATA_signal     279. .291
                     /gene="UGT1"
                     /note="Ritter,J.K., et al., 1992, J. Biol. Chem.,
                     267:3257-3261"
     mRNA            314. .>541
                     /gene="UGT1"
                     /product="bilirubin UDP-glucuronosyltransferase 1-1"
     CDS             330. .>541
                     /gene="UGT1"
                     /note="UDP glycosyltransferase 1"
                     /codon_start=1
                     /product="bilirubin UDP-glucuronosyltransferase 1-1"
                     /protein_id="AAF01205.1"
                     /db_xref="GI:6010650"
                     /translation="MAVESQGGRPLVLGLLLCVLGPVVSHAGKILLIPVDGSHWLSML
                     GAIQQLQQRGHEIVVLAPDASLYIRDG"
BASE COUNT      123 a    124 c    141 g    153 t
ORIGIN


  Query Match           100.0%;  Score 19;  DB 89;  Length 541;
  Best Local Similarity  100.0%;  Pred. No. 14;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tttgctcctgccagaggtt 19
          ||||||||||||||||||||
Db    323 TTTGCTCCTGCCAGAGGTT 305
```

```
RESULT  12
A65504/c
LOCUS       A65504        620 bp    DNA             PAT       29-MAR-1999
DEFINITION  Sequence 5 from Patent WO9732042.
ACCESSION   A65504
VERSION     A65504.1  GI:4531239
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 620)
  AUTHORS   Burchell,B.
  TITLE     DRUG TRIAL ASSAY SYSTEM
  JOURNAL   Patent: WO 9732042-A 5 04-SEP-1997;
            UNIV DUNDEE (GB)
COMMENT     Other publication AU 2224197 19970916.
FEATURES             Location/Qualifiers
     source          1. .620
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      157 a    127 c    151 g    185 t
ORIGIN


  Query Match           100.0%;  Score 19;  DB 9;  Length 620;
  Best Local Similarity  100.0%;  Pred. No. 14;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 tttgctcctgccagaggtt 19
          ||||||||||||||||||||
Db    605 TTTGCTCCTGCCAGAGGTT 587


RESULT  13
AF110194/c
LOCUS       AF110194      918 bp    DNA             PRI       02-JAN-2001
DEFINITION  Homo sapiens chromosome 2 UDP-glucuronosyltransferase (UGT1A1)
            gene, UGT1A1*33 allele, partial cds.
ACCESSION   AF110194
VERSION     AF110194.1  GI:12002134
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 918)
  AUTHORS   Guillemette,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-DEC-1998) Center for Cancer Research, MIT, 77
            Massachusetts Avenue, E17-540, Cambridge, MA 02139, USA
FEATURES             Location/Qualifiers
     source          1. .918
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="2"
     mRNA            <38. .>918
                     /gene="UGT1A1"
                     /product="UDP-glucuronosyltransferase"
     gene            <38...>918
                     /gene="UGT1A1"
                     /note="22. .918; 1A1 variant allele; L233R"
                     /allele="UGT1A1*33"
     CDS             38. .>918
                     /gene="UGT1A1"
                     /codon_start=1
                     /product="UDP-glucuronosyltransferase"
                     /protein_id="AAG43197.1"
                     /db_xref="GI:12002135"
                     /translation="MAVESQGGRPLVLGLLLCVLGPVVSHAGKILLIPVDGSHWLSML
                     GAIQQLQQRGHEIVVLAPDASLYIRDGAFYTLKTYPVPFQREDVKESFVSLGHNVFEN
                     DSFLQRVIKTYKKIKKDSAMLLSGCSHLLHNKELMASLAESSFDVMLTDPFLPCSPIV
                     AQYLSLPTVFFLHALPCSLEFEATQCPNPFSYVPRPLSSHSDHMTFLQRVKNMLIAFS
                     QNFLCDVVYSPYATRASEFLQREVTVQDLLSSASVWLFRSDFVKDYPRPIMPNMVFVG
```

GINCLHQNPLSQVCIGVG"
BASE COUNT 196 a 245 c 235 g 242 t
ORIGIN

Query Match 100.0%; Score 19; DB 88; Length 918;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tttgctcctgccagaggtt 19
||||||||||||||||||||
Db 31 TTTGCTCCTGCCAGAGGTT 13

RESULT 14
HUMUGT1A/c
LOCUS HUMUGT1A 1190 bp DNA PRI 14-JAN-1995
DEFINITION Human bilirubin UDP-glucuronosyltransferase (UGT1A) gene isozyme exon 1.
ACCESSION M84125
VERSION M84125.1 GI:340131
KEYWORDS bilirubin UDP-glucuronosyltransferase; isozyme.
SOURCE Homo sapiens (tissue library: cosmid) liver DNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1190)
AUTHORS Ritter,J.K., Crawford,J.M. and Owens,I.S.
TITLE Cloning of two human liver bilirubin UDP-glucuronosyltransferase cDNAs with expression in COS-1 cells
JOURNAL J. Biol. Chem. 266 (2), 1043-1047 (1991)
MEDLINE 91093210
REFERENCE 2 (bases 1 to 1190)
AUTHORS Ritter,J.K., Chen,F., Sheen,Y.Y., Tran,H.M., Kimura,S., Yeatman,M.T. and Owens,I.S.
TITLE A novel complex locus UGT1 encodes human bilirubin, phenol, and other UDP-glucuronosyltransferase isozymes with identical carboxyl termini
JOURNAL J. Biol. Chem. 267 (5), 3257-3261 (1992)
MEDLINE 92147680
FEATURES Location/Qualifiers
source 1. .1190
/organism="Homo sapiens"
/db_xref="taxon:9606"
/tissue_type="liver"
/tissue_lib="cosmid"
/lab_host="XL1-Blue"
gene 32. .46
/gene="UTG1A"
TATA_signal 32. .46
/gene="UTG1A"
exon 69. .948
/gene="UGT1A"
/number=1
gene 69. .1190
/gene="UGT1A"
CDS 85. .>948
/gene="UGT1A"
/codon_start=1
/product="bilirubin UDP-glucuronosyltransferase"
/protein_id="AAA61248.1"
/db_xref="GI:340132"
/translation="MAVESQGGRPLVLGLLLCVLGPVVSHAGKILLIPVDGSHWLSML
GAIQQLQQRGHEIVVLAPDASLYIRDGAFYTLKTYPVPFQREDVKESFVSLGHNVFEN
DSFLQRVIKTYKKIKKDSAMLLSGCSHLLHNKELMASLAESSFDVMLTDPFLPCSPIV
AQYLSLPTVFFLHALPCSLEFEATQCPNPFSYVPRPLSSHSDHMTFLQRVKNMLIAFS
QNFLCDVVYSPYATLASEFLQREVTVQDLLSSASVWLFRSDFVKDYPRPIMPNMVFVG
GINCLHQNPLSQ"
intron 949. .1190
/partial
/gene="UGT1A"
/note="does not fit consensus"
/number=1

BASE COUNT 261 a 290 c 286 g 353 t
ORIGIN

Query Match 100.0%; Score 19; DB 97; Length 1190;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tttgctcctgccagaggtt 19
||||||||||||||||||||
Db 78 TTTGCTCCTGCCAGAGGTT 60

RESULT 15
D87674/c
LOCUS D87674 3341 bp DNA PRI 14-APR-2000
DEFINITION Homo sapiens gene for bilirubin UDP-glucuronosyltransferase 1, promoter region and partial cds.
ACCESSION D87674
VERSION D87674.1 GI:3059176
KEYWORDS bilirubin UDP-glucuronosyltransferase 1.
SOURCE Homo sapiens DNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (sites)
AUTHORS Ueyama,H., Koiwai,O., Soeda,Y., Sato,H., Satoh,Y., Ohkubo,I. and Doida,Y.
TITLE Analysis of the promoter of human bilirubin UDP-glucuronosyltransferase gene (UGT1*1) in relevance to Gilbert's syndrome
JOURNAL Hepatol. Res. 9, 152-163 (1997)
REFERENCE 2 (bases 1 to 3341)
AUTHORS Ueyama,H.
TITLE Direct Submission
JOURNAL Submitted (04-SEP-1996) to the DDBJ/EMBL/GenBank databases. Hisao Ueyama, Shiga University of Medical Science, Department of Medical Biochemistry; Seta, Otsu, Shiga 520-21, Japan (Tel:077-548-2162, Fax:077-548-2164)
COMMENT Sequence updated (08-Jan-1997) by: Hisao Ueyama.
FEATURES Location/Qualifiers
source 1. .3341
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2q37"
misc_feature 2538. .2544
/note="XRE"
misc_feature 2606. .2610
/note="XRE"
misc_feature 3088. .3097
/note="E-box"
misc_feature 3101. .3113
/note="HNF-1 site"
CAAT_signal 3125. .3129
TATA_signal 3149. .3153
exon 3177. .3341
/number=1
gene 3192. .3341
/gene="UGT1*1"
CDS 3192. .>3341
/gene="UGT1*1"
/codon_start=1
/product="bilirubin UDP-glucuronosyltransferase 1"
/protein_id="BAA25600.1"
/db_xref="GI:3059177"
/translation="MAVESQGGRPLVLGLLLCVLGPVVSHAGKILLIPVDGSHWLSML
GAIQQL"
BASE COUNT 893 a 695 c 803 g 950 t
ORIGIN

Query Match 100.0%; Score 19; DB 91; Length 3341;
Best Local Similarity 100.0%; Pred. No. 13;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 tttgctcctgccagaggtt 19
           ||||||||||||||||||
Db    3185 TTTGCTCCTGCCAGAGGTT 3167
```

Search completed: July 25, 2001, 05:17:10
Job time: 9233 sec

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: July 25, 2001, 05:23:05 ; Search time 247.87 Seconds
(without alignments)
48.131 Million cell updates/sec

Title: US-09-142-095-4
Perfect score: 19
Sequence: 1 tttgctcctgccagaggtt 19

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_0601:*
1: /SIDS1/gcgdata/geneseq/geneseqn/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		19	100.0	19	18	AAT79544	UGT1*1 gene exon 1
c 2		19	100.0	620	18	AAT79540	Upstream DNA seque
c 3		18	94.7	1167	13	AAQ33024	UGT1A Exon 1 from
c 4		16	84.2	17	21	AAZ45058	Forward PCR primer
c 5		16	84.2	17	21	AAZ45074	Forward PCR primer
c 6		16	84.2	1307	22	AAF75200	Human 2hdeltalp DN
c 7		15.8	83.2	457	20	AAZ16518	Human gene express
c 8		15.8	83.2	843	19	AAV22955	cDNA encoding huma
c 9		15.8	83.2	1095	20	AAV68666	Nucleotide sequenc
c 10		15.8	83.2	1156	20	AAX31924	Human nodal protei
c 11		15.8	83.2	1571	22	AAF27725	Human transport pr
c 12		15.8	83.2	1945	21	AAC98907	Human pancreatic c
c 13		15.8	83.2	2003	19	AAV22956	Nucleotide sequenc
14		15.8	83.2	2818	17	AAT10554	Sheep mammary glan
15		15.8	83.2	3346	21	AAC47622	Arabidopsis thalia
c 16		14.8	77.9	239	21	AAF10976	Fusarium venenatum
c 17		14.8	77.9	359	21	AAC57066	Pinus radiata tran
c 18		14.8	77.9	375	21	AAC57038	Pinus radiata tran
19		14.8	77.9	386	21	AAC01025	Human secreted pro
20		14.8	77.9	404	21	AAC03243	Human secreted pro
21		14.8	77.9	405	20	AAX40904	Human secreted pro
c 22		14.8	77.9	430	21	AAC57046	Pinus radiata tran
c 23		14.8	77.9	469	21	AAC57049	Pinus radiata tran
c 24		14.8	77.9	471	21	AAC56976	Pinus radiata tran
25		14.8	77.9	479	18	AAT88072	Partial cDNA clone
26		14.8	77.9	479	18	AAV02162	Human secreted pro
c 27		14.8	77.9	508	21	AAZ80032	Human colon cancer
c 28		14.8	77.9	530	21	AAA57279	Human breast cance
29		14.8	77.9	697	20	AAZ33534	Human prostate can
c 30		14.8	77.9	800	21	AAC99839	Human secreted pro
31		14.8	77.9	1485	19	AAV18613	Nucleotide sequenc
32		14.8	77.9	1491	19	AAV18614	Nucleotide sequenc
33		14.8	77.9	2269	21	AAC67774	Human angiopoietin
34		14.8	77.9	2269	21	AAZ92213	Human angiopoietin
35		14.8	77.9	2269	21	AAZ92215	Human angiopoietin
36		14.8	77.9	2282	17	AAT44321	Human TIE-2 ligand
37		14.8	77.9	2282	17	AAT14650	Human TIE-2 ligand
38		14.8	77.9	2282	19	AAV18619	Human TIE-2 ligand
39		14.8	77.9	3496	19	AAV32554	Candida albicans C
c 40		14.8	77.9	4500	21	AAZ87998	Nucleotide sequenc
c 41		14.8	77.9	4590	21	AAC75767	Human ORFX ORF1322
42		14.8	77.9	5924	15	AAQ77790	Cobra CVF1 coding
43		14.8	77.9	5948	20	AAX85766	cDNA encoding cobr
c 44		14.8	77.9	45613	22	AAF28535	Genomic fragment #
c 45		14.8	77.9	143068	21	AAF21105	Human low adenosin

ALIGNMENTS

RESULT 1
AAT79544
ID AAT79544 standard; DNA; 19 BP.
XX
AC AAT79544;
XX
DT 23-JAN-1998 (first entry)
XX
DE UGT1*1 gene exon 1 upstream PCR primer D.
XX
KW Uridine diphosphate glucuronosyltransferase gene; UGT;
KW Gilbert's syndrome; GS; unconjugated hyperbilirubinaemia;
KW bilirubin glucuronidation; Crigler-Najjar; type 2; drug metabolism;
KW Drug trial efficiency; screening; PCR primer; ss.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN WO9732042-A2.
XX
PD 04-SEP-1997.
XX
PF 03-MAR-1997; 97WO-GB00577.
XX
PR 16-MAR-1996; 96GB-0005598.
PR 01-MAR-1996; 96GB-0004480.
XX
PA (UYDU-) UNIV DUNDEE.
XX
PI Burchell B;
XX
DR WPI; 1997-448702/41.
XX
PT Improving drug trial efficiency comprises identifying participants

```
PT   with Gilbert's syndrome - useful as their altered drug metabolism
PT   may hinder result interpretation
XX
PS   Claim 14; Page 12; 31pp; English.
XX
CC   This PCR primer (with primers AAT79542-44) flanks the TATA box sequence
CC   upstream of the uridine diphosphate glucuronosyltransferase (UGT) gene
CC   1*1 exon 1 (see AAT79540), and was used to amplify fragments of 98-100
CC   bp. This gene is known to be associated with Gilbert's syndrome (GS). GS
CC   is a mild, common form of unconjugated hyperbilirubinaemia associated
CC   with reduced bilirubin glucuronidation capacity. Analysis of the genetic
CC   basis of GS has allowed 2 forms to be identified. One is a mild form
CC   associated with a homozygous 2 bp insertion in the TATA sequence
CC   upstream of the UGT1*1 exon 1, and the other is a more severe form
CC   associated with heterozygosity for a mutation which, when homozygous,
CC   causes Crigler-Najjar type 2 disease. The first form is autosomal
CC   recessive and the second is inherited dominantly. Patients suffering from
CC   GS, which is benign, may have altered metabolism of some drugs, making it
CC   difficult to determine if an effect is due to the drug or the syndrome.
CC   Drug trial efficiency would be improved if potential participants can be
CC   screened for the genetic basis of GS, and eliminated or included  on
CC   basis of them possessing or not possessing GS.
XX
SQ   Sequence 19 BP; 2 A; 5 C; 5 G; 7 T; 0 other;


  Query Match             100.0%;  Score 19;  DB 18;  Length 19;
  Best Local Similarity   100.0%;  Pred. No. 1.5;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 tttgctcctgccagaggtt 19
           ||||||||||||||||||||
Db       1 tttgctcctgccagaggtt 19


RESULT   2
AAT79540/c
ID   AAT79540 standard; DNA; 620 BP.
XX
AC   AAT79540;
XX
DT   23-JAN-1998  (first entry)
XX
DE   Upstream DNA sequence of UGT1*1 gene exon 1.
XX
KW   Uridine diphosphate glucuronosyltransferase gene; UGT;
KW   Gilbert's syndrome; GS; unconjugated hyperbilirubinaemia;
KW   bilirubin glucuronidation; Crigler-Najjar; type 2; drug metabolism;
KW   Drug trial efficiency; screening; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   protein_bind     2..8
FT                    /*tag=  a
FT                    /bound_moeity= AP1
FT   GC_signal        93..102
FT                    /*tag=  b
FT                    /note= "SP1 binding site"
FT   misc_feature     180..191
FT                    /*tag=  c
FT                    /note= "feature indicated in patent, but no further
FT                          explanation is given"
FT   protein_bind     316..323
FT                    /*tag=  d
FT                    /bound_moeity= AP3
FT   protein_bind     317..324
FT                    /*tag=  e
FT                    /bound_moeity= CLBP
FT   misc_feature     366..376
FT                    /*tag=  f
FT                    /note= "feature indicated in patent, but no further
FT                          explanation is given"
FT   protein_bind     386..392
FT                    /*tag=  g
FT                    /bound_moeity= AP1
FT   protein_bind     513..519
FT                    /*tag=  h
FT                    /bound_moeity= AP1
FT   misc_feature     513..520
FT                    /*tag=  i
FT                    /note= "feature indicated in patent, but no further
FT                          explanation is given"
FT   protein_bind     520..532
FT                    /*tag=  j
FT                    /bound_moeity= HNF1
FT   TATA_signal      558..572
FT                    /*tag=  k
FT                    /note= "corresponds to positions -53 to -39 in patent"
FT   misc_feature     595
FT                    /*tag=  l
FT                    /note= "feature indicated in patent, but no further
FT                          explanation is given"
FT   CDS              612
FT                    /*tag=  m
FT                    /codon_start= 612
XX
PN   WO9732042-A2.
XX
PD   04-SEP-1997.
XX
PF   03-MAR-1997;   97WO-GB00577.
XX
PR   16-MAR-1996;   96GB-0005598.
PR   01-MAR-1996;   96GB-0004480.
XX
PA   (UYDU-) UNIV DUNDEE.
XX
PI   Burchell B;
XX
DR   WPI; 1997-448702/41.
XX
PT   Improving drug trial efficiency comprises identifying participants
PT   with Gilbert's syndrome - useful as their altered drug metabolism
PT   may hinder result interpretation
XX
PS   Claim 11; Fig 4; 31pp; English.
XX
CC   This sequence represents the upstream sequence, positions -611 to 9
CC   in the patent, of uridine diphosphate glucuronosyltransferase (UGT)
CC   gene 1*1 exon 1. This gene is known to be associated with Gilbert's
CC   syndrome (GS). GS is a mild, common form of unconjugated
CC   hyperbilirubinaemia associated with reduced bilirubin glucuronidation
CC   capacity. Analysis of the genetic basis of GS has allowed 2 forms to be
CC   identified. One is a mild form associated with a homozygous 2 bp
CC   insertion in the TATA sequence upstream of the UGT1*1 exon 1, and
CC   the other is a more severe form associated with heterozygosity for a
CC   mutation which, when homozygous, causes Crigler-Najjar type 2 disease.
CC   The first form is autosomal recessive and the second is inherited
CC   dominantly. Patients suffering from GS, which is benign, may have
CC   altered metabolism of some drugs, making it difficult to determine if an
CC   effect is due to the drug or the syndrome. Drug trial efficiency would
CC   be improved if potential participants can be screened for the genetic
CC   basis of GS, and eliminated or included  on basis of them possessing or
CC   not possessing GS. In this case, screening involves PCR amplification
CC   of the UGT gene, using the primers described in AAT79541-44.
XX
SQ   Sequence 620 BP; 157 A; 127 C; 151 G; 185 T; 0 other;


  Query Match             100.0%;  Score 19;  DB 18;  Length 620;
  Best Local Similarity   100.0%;  Pred. No. 2.2;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 tttgctcctgccagaggtt 19
           ||||||||||||||||||||
```

```
Db    605 TTTGCTCCTGCCAGAGGTT 587


RESULT   3
AAQ33024/c
ID   AAQ33024 standard; DNA; 1167 BP.
XX
AC   AAQ33024;
XX
DT   27-JAN-1993  (first entry)
XX
DE   UGT1A Exon 1 from the UGT1 gene locus.
XX
KW   UGT1A; UGT1BP; UGT1C; UGT1D; UGT1E; UGT1F; isozyme; bilirubin;
KW   UDP-glucuronosyl transferase; CN; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   misc_RNA        1..20
FT                   /*tag=  a
FT                   /note= "representation of 11.7 kbp of
FT                          non-sequenced DNA between the sequences
FT                          represented in AAQ33023 and AAQ33024"
FT   TATA_signal     52..66
FT                   /*tag=  b
FT   exon            166..909
FT                   /*tag=  c
FT                   /label= UGT1A_Exon_1
FT   misc_RNA        181..909
FT                   /*tag=  d
FT                   /note= "encodes transferase isoform; see CC"
FT   misc_RNA        1148..1167
FT                   /*tag=  e
FT                   /note= "representation of 5 kbp of
FT                          non-sequenced DNA between the sequences
FT                          represented in AAQ33024 and AAQ33025"
XX
PN   WO9212987-A.
XX
PD   06-AUG-1992.
XX
PF   10-JAN-1992;   92WO-US00282.
XX
PR   10-JAN-1991;   91US-0639453.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICE.
XX
PI   Owens IS,  Ritter JK;
XX
DR   WPI; 1992-284593/34.
DR   P-PSDB; AAR30194.
XX
PT   Isolated gene locus UGT1, DNA segments and diagnostic probes -
PT   for diagnosing Gilbert's disease and Crigler-Najjar syndrome
PT   types I and II
XX
PS   Disclosure; Fig 1F; 99pp; English.
XX
CC   The isolated gene locus, UGT1, has a sequence of about 10000 bp
CC   which represent (1) Exon 1, comprising 6 transcriptional units
CC   (UGT1F, E, D, C, BP and A), represented in AAQ27368 and
CC   AAQ33020-24 respectively;
CC                  (2) Exon 2, represented in AAQ33025;
CC                  (3) Exon 3, represented in AAQ33026;
CC                  (4) Exon 4, represented in AAQ33026;
CC                  (5) Exon 5, represented in AAQ33027; and
CC                  (6) about 69 kb of non-sequenced DNA.
CC   Six unique N-termini of 286-289 amino acids are encoded by
CC   the six different first exons and identical C-termini of 246 amino
CC   acids are encoded by the common exons 2-5.  The UGT1 gene locus
CC   encodes a family of UDP-glucuronosyl transferase isozymes, two of
CC   which metabolise bilirubin.
CC   Patients having Crigler-Najjar Syndrome (CN) Type I, have a
CC   mutation present in the second common exon.
XX
SQ   Sequence 1167 BP; 255 A; 259 C; 272 G; 340 T; 41 other;


  Query Match             94.7%;  Score 18;  DB 13;  Length 1167;
  Best Local Similarity   100.0%;  Pred. No. 7.5;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 ttgctcctgccagaggtt 19
          ||||||||||||||||||
Db    174 TTGCTCCTGCCAGAGGTT 157


RESULT   4
AAZ45058/c
ID   AAZ45058 standard; DNA; 17 BP.
XX
AC   AAZ45058;
XX
DT   28-FEB-2000  (first entry)
XX
DE   Forward PCR primer used in the secondary amplification of UGT1 exon 1A.
XX
KW   Uridine diphosphate-glucuronosyltransferase 1; UGT1; polymorphism; probe;
KW   glucuronic acid; Crigler-Najjar syndrome; Gilbert syndrome; jaundice;
KW   unconjugated hyperbilirubinaemia; drug metabolism; transgenic animal;
KW   pharmacogenetic screening; diagnose; PCR primer; ss.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   WO9957322-A2.
XX
PD   11-NOV-1999.
XX
PF   04-MAY-1999;   99WO-US09702.
XX
PR   07-MAY-1998;   98US-0084807.
XX
PA   (AXYS-) AXYS PHARM INC.
XX
PI   Penny L,  Galvin M;
XX
DR   WPI; 2000-052981/04.
XX
PT   New nucleic acid representing polymorphisms in the human uridine
PT   diphosphate glucuronosyltransferase gene, used for diagnosis and
PT   evaluation of drug metabolism  -
XX
PS   Examples; Page 16; 63pp; English.
XX
CC   PCR primers AAZ45042-Z45073 are used to amplify human uridine
CC   diphosphate-glucuronosyltransferase 1 (UGT1) exon sequences. The UGTs
CC   are a family of enzymes that catalyse the glucuronic acid conjugation of
CC   a wide range of endogenous and exogenous substrates including  phenols,
CC   alcohols, amines and fatty acids. Many of the reactions catalysed by
CC   UGTs result in toxic substances being converted to compounds which are
CC   more water soluble and are excreted. The invention relates to and
CC   identifies UGT1 polymorphisms (AAZ45004-Z45041). The polymorphism
CC   sequences are useful as probes for detecting UGT1 locus polymorphisms,
CC   indicative of altered UGT1 expression or activity. These polymorphisms
CC   are associated with Crigler-Najjar and Gilbert syndromes (unconjugated
CC   hyperbilirubinaemia) and drug metabolism. The genotyping of the UGT1 gene
CC   is used to predict the rate of metabolism of UGT1 substrates, possible
CC   drug-drug interactions and adverse side effects (i.e. to optimize drug
CC   dosage), and to screen for diseases caused by exposure to toxins and to
CC   study the effects of polymorphisms on enzymatic activity. The UGT1
CC   sequences, including polymorphisms, can also be used to produce the
CC   corresponding protein (or its fragments) or to generate transgenic
CC   animals or modified cells e.g. for pharmacogenetic screening.
XX
```

```
SQ   Sequence 17 BP; 5 A; 4 C; 6 G; 2 T; 0 other;

  Query Match             84.2%;  Score 16;  DB 21;  Length 17;
  Best Local Similarity   100.0%;  Pred. No. 45;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 tttgctcctgccagag 16
           ||||||||||||||||
Db      16 TTTGCTCCTGCCAGAG 1


RESULT   5
AAZ45074/c
ID   AAZ45074 standard; DNA; 17 BP.
XX
AC   AAZ45074;
XX
DT   28-FEB-2000  (first entry)
XX
DE   Forward PCR primer for sequencing UGT1 exon 1A polymorphism #1.
XX
KW   Uridine diphosphate-glucuronosyltransferase 1; UGT1; polymorphism; probe;
KW   glucuronic acid; Crigler-Najjar syndrome; Gilbert syndrome; jaundice;
KW   unconjugated hyperbilirubinaemia; drug metabolism; transgenic animal;
KW   pharmacogenetic screening; diagnose; PCR primer; ss.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   WO9957322-A2.
XX
PD   11-NOV-1999.
XX
PF   04-MAY-1999;   99WO-US09702.
XX
PR   07-MAY-1998;   98US-0084807.
XX
PA   (AXYS-) AXYS PHARM INC.
XX
PI   Penny L,  Galvin M;
XX
DR   WPI; 2000-052981/04.
XX
PT   New nucleic acid representing polymorphisms in the human uridine
PT   diphosphate glucuronosyltransferase gene, used for diagnosis and
PT   evaluation of drug metabolism -
XX
PS   Examples; Page 19; 63pp; English.
XX
CC   Primers AAZ45074-Z45109 are used to sequence the human uridine
CC   diphosphate-glucuronosyltransferase 1 (UGT1) exon polymorphism
CC   sequences. The UGTs are a family of enzymes that catalyse the glucuronic
CC   acid conjugation of a wide range of endogenous and exogenous substrates
CC   including  phenols, alcohols, amines and fatty acids. Many of the
CC   reactions catalysed by UGTs result in toxic substances being converted
CC   to compounds which are more water soluble and are excreted. The
CC   invention relates to and identifies UGT1 polymorphisms (AAZ45004-Z45041).
CC   The polymorphism sequences are useful as probes for detecting UGT1 locus
CC   polymorphisms, indicative of altered UGT1 expression or activity. These
CC   polymorphisms are associated with Crigler-Najjar and Gilbert syndromes
CC   (unconjugated hyperbilirubinaemia) and drug metabolism. The genotyping
CC   of the UGT1 gene is used to predict the rate of metabolism of UGT1
CC   substrates, possible drug-drug interactions and adverse side effects
CC   (i.e. to optimize drug dosage), and to screen for diseases caused by
CC   exposure to toxins and to study the effects of polymorphisms on
CC   enzymatic activity. The UGT1 sequences, including polymorphisms, can also
CC   be used to produce the corresponding protein (or its fragments) or to
CC   generate transgenic animals or modified cells e.g. for pharmacogenetic
CC   screening.
XX
SQ   Sequence 17 BP; 5 A; 4 C; 6 G; 2 T; 0 other;

  Query Match             84.2%;  Score 16;  DB 21;  Length 17;
  Best Local Similarity   100.0%;  Pred. No. 45;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 tttgctcctgccagag 16
           ||||||||||||||||
Db      16 TTTGCTCCTGCCAGAG 1


RESULT   6
AAF75200/c
ID   AAF75200 standard; DNA; 1307 BP.
XX
AC   AAF75200;
XX
DT   15-MAY-2001  (first entry)
XX
DE   Human 2hdelta1p DNA.
XX
KW   Notch receptor; ligand; cancer; melanoma; ischemia; ds.
XX
OS   Homo sapiens.
XX
PN   WO200112664-A2.
XX
PD   22-FEB-2001.
XX
PF   17-AUG-2000; 2000WO-US22609.
XX
PR   19-AUG-1999;   99US-0149934.
XX
PA   (CHIR ) CHIRON CORP.
XX
PI   Vivien C,  Rohan M,  Williams LT;
XX
DR   WPI; 2001-211201/21.
XX
PT   Novel notch receptor ligands useful for modulating angiogenesis and
PT   immune responses for treating rheumatoid arthritis, cancer-related
PT   angiogenesis to stop tumor growth, and as diagnostic reagents -
XX
PS   Disclosure; Fig 6; 66pp; English.
XX
CC   The present invention relates to a novel Notch receptor ligand.
CC   The invention is useful for detecting Notch ligand expression in
CC   human cancer cells or melanoma cells. Also useful for enhancing
CC   angiogenesis in a mammal, useful when the mammal exhibits
CC   tissue ischemia.
XX
SQ   Sequence 1307 BP; 294 A; 328 C; 396 G; 289 T; 0 other;

  Query Match             84.2%;  Score 16;  DB 22;  Length 1307;
  Best Local Similarity   100.0%;  Pred. No. 75;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4 gctcctgccagaggtt 19
           ||||||||||||||||
Db    1119 GCTCCTGCCAGAGGTT 1104


RESULT   7
AAZ16518/c
ID   AAZ16518 standard; cDNA; 457 BP.
XX
AC   AAZ16518;
XX
DT   12-OCT-1999  (first entry)
XX
DE   Human gene expression product cDNA sequence SEQ ID NO:3988.
XX
KW   Human; gene; gene expression product; diagnosis; therapy; probe;
```

```
KW   detection; mapping; tissue typing; profiling; forensic; cancer;
KW   genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.
XX
OS   Homo sapiens.
XX
PN   WO9938972-A2.
XX
PD   05-AUG-1999.
XX
PF   28-JAN-1999;   99WO-US01619.
XX
PR   03-APR-1998;   98US-0080666.
PR   28-JAN-1998;   98US-0072910.
PR   24-FEB-1998;   98US-0075954.
PR   31-MAR-1998;   98US-0080114.
PR   03-APR-1998;   98US-0080515.
XX
PA   (CHIR ) CHIRON CORP.
PA   (HYSE-) HYSEQ INC.
XX
PI   Crkvenjakov R, Dickson M, Drmanac R, Drmanac S;
PI   Escobedo J, Garcia PD, Garcia V, Giese K, Innis MA;
PI   Jones WL, Kassam A, Kennedy GC, Kita D, Labat I;
PI   Lamson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;
PI   Stache-Crain B, Sudduth-Klinger J, Williams LT;
XX
DR   WPI; 1999-494092/41.
XX
PT   Novel human genes and their expression products which are
PT   differentially expressed in different cell types
XX
PS   Claim 1; Page 1892-1893; 2479pp; English.
XX
CC   The present invention describes a library of human polynucleotides
CC   comprising the sequences given in AAZ12532 to AAZ17779. Also described is
CC   a method of detecting differentially expressed genes correlated with the
CC   cancerous state of a mammalian cell, comprising detecting at least one
CC   differentially expressed gene product in a test sample from a cell
CC   suspected of being cancerous, where the gene product is encoded by one
CC   of the 5248 polynucleotide sequences given in AAZ12532 to AAZ17779. The
CC   polynucleotides can be used as a source of primers and probes, which can
CC   be used for a variety of purpose, e.g. detection of expression levels,
CC   mapping, tissue typing or profiling, forensics, genetic analysis and
CC   detection of polymorphisms. Polypeptides encoded by the polynucleotides
CC   can be used for raising antibodies for experimental, diagnostic and
CC   therapeutic purposes. The polynucleotides may also be used to construct
CC   arrays for diagnostics (which may be used to determine function of an
CC   encoded protein); and to detect differences in expression levels between
CC   two cells (e.g. to identify abnormal or diseased tissue in a human, to
CC   identify a genetic predisposition or susceptibility to a disease such as
CC   cancer). The polynucleotides of the invention are especially used in the
CC   diagnosis, prognosis and management of colorectal cancer, breast cancer,
CC   and lung cancer. The polynucleotides can also be used to screen for
CC   peptide analogues and antagonists.
XX
SQ   Sequence 457 BP; 120 A; 88 C; 93 G; 139 T; 17 other;


  Query Match             83.2%;  Score 15.8;  DB 20;  Length 457;
  Best Local Similarity   89.5%;  Pred. No. 83;
  Matches   17;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 tttgctcctgccagaggtt 19
           ||||||||| |||||| ||
Db     300 TTTGCTCCTCCCAGAGCTT 282


RESULT   8
AAV22955/c
ID   AAV22955 standard; cDNA; 843 BP.
XX
AC   AAV22955;
XX
```

```
DT   04-AUG-1998  (first entry)
XX
DE   cDNA encoding human bone morphogenetic protein-16 (BMP-16).
XX
KW   Human; bone morphogenetic protein-16; BMP-16; murine protein; nodal;
KW   formation; bone; cartilage; treatment; wound healing; reduction;
KW   fibrosis; scar tissue formation; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..843
FT                   /*tag= a
FT   sig_peptide     1..510
FT                   /*tag= b
FT   mat_peptide     511..840
FT                   /*tag= c
XX
PN   WO9812322-A1.
XX
PD   26-MAR-1998.
XX
PF   09-JUL-1997;   97WO-US11954.
XX
PR   18-SEP-1996;   96US-0715202.
XX
PA   (GEMY ) GENETICS INST INC.
XX
PI   Celeste AJ, Murray BL;
XX
DR   WPI; 1998-217262/19.
DR   P-PSDB; AAW56477.
XX
PT   New isolated bone morphogenetic protein-16 - used to develop
PT   products for inducing formation of bone, cartilage and other
PT   connective tissue, particularly for wound healing and tissue repair
XX
PS   Claim 5; Pages 33-34; 43pp; English.
XX
CC   The present sequence encodes a human bone morphogenetic protein-16
CC   (BMP-16). Human BMP-16 is a homologue of a murine protein celled nodal,
CC   which is expressed in the mouse node during gastrulation. BMP-16 cDNA is
CC   isolated from a human genomic library screened with a probe derived from
CC   the nodal DNA sequence. The BMP-16 proteins can induce the formation of
CC   bone, cartilage or other connective tissue. They can be used for treating
CC   bone, cartilage or other connective tissue defects, periodontal disease
CC   or healing of various types of tissues and wounds. They can also increase
CC   neuronal, astrocytic and glial cell survival and therefore be useful in
CC   transplantation and treatment of conditions exhibiting a decrease in
CC   neuronal survival and repair. They can also exhibit properties such as
CC   angiogenic, chemotactic and/or chemoattractant properties, and effects
CC   on cells including induction of collagen synthesis, fibrosis,
CC   differentiation responses, cell proliferative responses and responses
CC   involving cell adhesion, migration and extracellular matrices. These
CC   properties make the proteins potential agents for wound  healing,
CC   reduction of fibrosis and reduction of scar tissue formation.
XX
SQ   Sequence 843 BP; 176 A; 241 C; 244 G; 182 T; 0 other;


  Query Match             83.2%;  Score 15.8;  DB 19;  Length 843;
  Best Local Similarity   89.5%;  Pred. No. 89;
  Matches   17;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 tttgctcctgccagaggtt 19
           | ||||||||| |||||||
Db     412 TCTGCTCCTGCGAGAGGTT 394


RESULT   9
AAV68666/c
ID   AAV68666 standard; DNA; 1095 BP.
XX
```

```
AC   AAV68666;
XX
DT   03-JUN-1999  (first entry)
XX
DE   Nucleotide sequence of the human Tango-78 gene.
XX
KW   Human; Tango-78; host cell; recombinant protein; antibody;
KW   receptor; specific binding agent; probe; primer; hybridisation;
KW   amplification; mutation; genetic mapping; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             588..1094
FT                   /*tag= a
FT                   /product= "Tango-78"
XX
PN   WO9906427-A1.
XX
PD   11-FEB-1999.
XX
PF   04-AUG-1998;   98WO-US16241.
XX
PR   04-AUG-1997;   97US-0054645.
XX
PA   (MILL-) MILLENNIUM BIOTHERAPEUTICS INC.
XX
PI   McCarthy SA;
XX
DR   WPI; 1999-153693/13.
DR   P-PSDB; AAW84595.
XX
PT   New nucleic acid encoding human Tango-78, -79 and -81 proteins -
PT   useful for diagnosis and treatment of Tango-associated diseases
XX
PS   Claim 1; Fig 1; 67pp; English.
XX
CC   This is the nucleotide sequence encoding the human Tango-78
CC   protein used in the method of the invention. Host cells containing
CC   the Tango protein are used to produce recombinant proteins for
CC   raising antibodies. It is also used in identifying specific
CC   binding agents (including cognate receptors), which can be used to
CC   determine amounts of recombinant protein in cells or
CC   therapeutically. Antibodies or other specific binding agents, are
CC   used to detect recombinant proteins and fragments of the Tango
CC   nucleotide sequence can be used as probes or primers for detecting
CC   the Tango gene, specifically mRNA, in usual hybridisation or
CC   amplification assays. These assays are used for diagnosis of
CC   diseases associated with abnormal expression of Tango proteins,
CC   e.g. detecting mutations in the Tango gene. Fragments of the Tango
CC   nucleic acid sequence are also used for genetic mapping and
CC   chromosome identification, and as antisense, ribozyme or
CC   triplex-forming therapeutics. Antibodies may also be used to generate
CC   anti-idiotype antibodies.
XX
SQ   Sequence 1095 BP; 231 A; 296 C; 312 G; 256 T; 0 other;


  Query Match             83.2%;  Score 15.8;  DB 20;  Length 1095;
  Best Local Similarity   89.5%;  Pred. No. 92;
  Matches   17;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 tttgctcctgccagaggtt 19
           | ||||||||| |||||||
Db     801 TCTGCTCCTGCGAGAGGTT 783


RESULT  10
AAX31924/c
ID   AAX31924 standard; DNA; 1156 BP.
XX
AC   AAX31924;
XX
```

```
DT   18-JUN-1999  (first entry)
XX
DE   Human nodal protein encoding DNA.
XX
KW   Nodal protein; lefty protein; TGF-beta; sexual development; human; bone;
KW   pituitary; cartilage; osteoarthritis; osteoporosis; haematopoiesis;
KW   periodontal disease; wound healing; tissue repair; tumour; cancer;
KW   interstitial lung disease; autoimmunity; leukaemia; lymphoma; immunity;
KW   immunosuppression; inflammatory bowel disease; myelosuppression;
KW   infectious disease; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..852
FT                   /*tag= a
FT                   /product= "Nodal protein"
FT                   /note= "the start codon is not indicated"
FT   misc_feature    517..849
FT                   /*tag= b
FT                   /note= "sequence coding for the active fragment of
FT                          the Nodal polypeptide"
XX
PN   WO9909198-A1.
XX
PD   25-FEB-1999.
XX
PF   20-AUG-1998;   98WO-US17211.
XX
PR   21-AUG-1997;   97US-0056565.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Ebner R,  Ruben SM,  Soppet DR;
XX
DR   WPI; 1999-190173/16.
DR   P-PSDB; AAY03849.
XX
PT   New isolate human Nodal and Lefty polypeptides
XX
PS   Claim 2; Fig 1A; 182pp; English.
XX
CC   The present invention relates to novel human nodal and lefty proteins
CC   which are members of the TGF-beta family. The human nodal and lefty
CC   proteins may be involved in a developmental process such as the correct
CC   formation of various structures or in one or more post-developmental
CC   capacities including sexual development, pituitary hormone production,
CC   and the creation of bone and cartilage. The Nodal and Lefty polypeptides
CC   are useful for enhancing or enriching the growth and/or differentiation
CC   of specific cell populations, eg. embryonic cells or stem cells. They can
CC   be used to treat such conditions as osteoarthritis, osteoporosis, and
CC   other abnormalities of bone, cartilate, muscle, tendon, ligament, and/or
CC   other connective tissues and/or organs such as liver, lung, cardiac,
CC   pancreas, and kidney. Compositions containing nodal and lefty proteins
CC   may be useful for growth formation, for treating periodontal disease and
CC   for modulating haematopoiesis, wound healing and tissue repair. They can
CC   also be used for the treatment of tumours, cancers, interstitial lung
CC   disease, and any disregulation of the growth and differentiation patterns
CC   of cell function including autoimmunity, arthritis, leukaemia, lymphomas,
CC   immunosuppression, immunity, humoral immunity, inflammatory bowel
CC   disease, myelosuppression, or infectious diseases. The present sequence
CC   represents a DNA encoding a human nodal polypeptide. The cDNA encoding
CC   the nodal protein is deposited under the ATCC deposit No. 209092 and/or
CC   209135.
XX
SQ   Sequence 1156 BP; 285 A; 298 C; 340 G; 233 T; 0 other;


  Query Match             83.2%;  Score 15.8;  DB 20;  Length 1156;
  Best Local Similarity   89.5%;  Pred. No. 93;
  Matches   17;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 tttgctcctgccagaggtt 19
```

```
          \ ||||||||| |||||||
Db    421 TCTGCTCCTGCGAGAGGTT 403


RESULT  11
AAF27725/c
ID   AAF27725 standard; cDNA; 1571 BP.
XX
AC   AAF27725;
XX
DT   28-MAR-2001  (first entry)
XX
DE   Human transport protein TPPT-25 coding sequence.
XX
KW   Human; transport protein; TPPT; transport disorder; metabolic disorder;
KW   neurological disorder; cardiovascular disorder; reproductive disorder;
KW   immune disorder; cancer; ss.
XX
OS   Homo sapiens.
XX
PN   WO200078953-A2.
XX
PD   28-DEC-2000.
XX
PF   16-JUN-2000; 2000WO-US16668.
XX
PR   17-JUN-1999;   99US-0139923.
PR   10-AUG-1999;   99US-0148177.
PR   18-AUG-1999;   99US-0149357.
PR   28-OCT-1999;   99US-0162287.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Lal P, Yang J, Yue H, Hillman JL, Tang YT, Bandman O, Burford N;
PI   Baughn MR, Azimzai Y, Lu DAM, Au-Young J, Patterson C;
XX
DR   WPI; 2001-041424/05.
DR   P-PSDB; AAB60105.
XX
PT   Isolated polypeptide with a human transport protein sequence is useful
PT   for the diagnosis, prevention and treatment of disorders associated
PT   with the immune, reproductive and cardiovascular systems -
XX
PS   Claim 5; Page 155-156; 165pp; English.
XX
CC   The present invention provides the protein and coding sequences for 43
CC   novel human transport proteins (designated TPPTs). These can be used in
CC   the diagnosis and treatment of transport, metabolic, neurological,
CC   reproductive, cardiovascular and immune disorders, and cell proliferative
CC   disorders such as cancer
XX
SQ   Sequence 1571 BP; 52[illegible]2 A; 265 C; 289 G; 495 T; 0 other;

  Query Match           83.2%;  Score 15.8;  DB 22;  Length 1571;
  Best Local Similarity 89.5%;  Pred. No. 96;
  Matches   17;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 tttgctcctgccagaggtt 19
          ||| ||| ||||||||||||
Db   1534 TTTACTCCTGCCAGAGGTT 1516


RESULT  12
AAC98907/c
ID   AAC98907 standard; cDNA; 1945 BP.
XX
AC   AAC98907;
XX
DT   09-MAR-2001  (first entry)
XX
DE   Human pancreatic cancer antigen nucleotide sequence SEQ ID NO:135.
XX
KW   Human; pancreas; pancreatic cancer; pancreatic cancer antigen;
KW   detection; diagnosis; identification; cytostatic; neuroprotective;
KW   nootropic; immunomodulatory; relaxant; contraceptive; gynaecological;
KW   antiinflammatory; cardiant; gene therapy; chromosome mapping;
KW   linkage analysis; tissue identification; tissue typing; forensic;
KW   neural; immune system; muscular; reproductive; gastrointestinal;
KW   pulmonary; cardiovascular; renal; proliferative; ss.
XX
OS   Homo sapiens.
XX
PN   WO200055320-A1.
XX
PD   21-SEP-2000.
XX
PF   08-MAR-2000; 2000WO-US05989.
XX
PR   12-MAR-1999;   99US-0124270.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA, Ruben SM;
XX
DR   WPI; 2000-579444/54.
DR   P-PSDB; AAB54142.
XX
PT   New nucleic acid that is a pancreatic cancer antigen for preventing,
PT   treating, or ameliorating a medical condition, particular pancreatic
PT   cancer, or for use in assays for diagnosing a pathological condition -
XX
PS   Claim 1; Page 594-595; 1379pp; English.
XX
CC   AAC98773 to AAC99231 encode the human pancreatic cancer associated
CC   proteins, called pancreatic cancer antigens, given in AAB54008 to
CC   AAB54466. The human pancreatic cancer antigens have cytostatic,
CC   neuroprotective, nootropic, immunomodulatory, relaxant, contraceptive,
CC   gynaecological, cardiant and antiinflammatory activities, and can be used
CC   in gene therapy. The polynucleotide and proteins can be used for
CC   preventing, treating, or ameliorating a medical condition or in assays
CC   for diagnosing a pathological condition or a susceptibility to one in a
CC   subject. Binding partners to the proteins and the activity of the
CC   proteins can be identified. The pancreatic cancer antigens can be used to
CC   detect, treat or prevent pancreatic disorders, especially cancer.
CC   Agonists and antagonists to the antigens can be screened for. The
CC   pancreatic cancer antigen polynucleotides can be used to design nucleic
CC   acid hybridisation probes that can be used in chromosome mapping, linkage
CC   analysis, tissue identification and/or typing and a variety of forensic
CC   and diagnostic methods. The proteins can be used to generate antibodies
CC   which are used to purify, detect and target the polypeptides, including
CC   both in vivo and in vitro diagnostic and therapeutic methods. The
CC   proteins can be used to treat or prevent neural, immune system, muscular,
CC   reproductive, gastrointestinal, pulmonary, cardiovascular, renal or
CC   proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent
CC   sequences used in the exemplification of the present invention.
XX
SQ   Sequence 1945 BP; 533 A; 3[illegible]0 C; 477 G; 552 T; 13 other;

  Query Match           83.2%;  Score 15.8;  DB 21;  Length 1945;
  Best Local Similarity 89.5%;  Pred. No. 98;
  Matches   17;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 tttgctcctgccagaggtt 19
          ||| |||||||||||||||
Db   1203 TTTACTCCTGCCAGAGCTT 1185


RESULT  13
AAV22956/c
ID   AAV22956 standard; DNA; 2003 BP.
XX
AC   AAV22956;
XX
DT   04-AUG-1998  (first entry)
```

PR	29-MAR-1999;	99US-0126785.
PR	01-APR-1999;	99US-0127462.
PR	06-APR-1999;	99US-0128234.
PR	08-APR-1999;	99US-0128714.
PR	16-APR-1999;	99US-0129845.
PR	19-APR-1999;	99US-0130077.
PR	21-APR-1999;	99US-0130449.
PR	23-APR-1999;	99US-0130510.
PR	23-APR-1999;	99US-0130891.
PR	28-APR-1999;	99US-0131449.
PR	30-APR-1999;	99US-0132048.
PR	30-APR-1999;	99US-0132407.
PR	04-MAY-1999;	99US-0132484.
PR	05-MAY-1999;	99US-0132485.
PR	06-MAY-1999;	99US-0132486.
PR	06-MAY-1999;	99US-0132487.
PR	07-MAY-1999;	99US-0132863.
PR	11-MAY-1999;	99US-0134256.
PR	14-MAY-1999;	99US-0134218.
PR	14-MAY-1999;	99US-0134219.
PR	14-MAY-1999;	99US-0134221.
PR	14-MAY-1999;	99US-0134370.
PR	18-MAY-1999;	99US-0134768.
PR	19-MAY-1999;	99US-0134941.
PR	20-MAY-1999;	99US-0135124.
PR	21-MAY-1999;	99US-0135353.
PR	24-MAY-1999;	99US-0135629.
PR	25-MAY-1999;	99US-0136021.
PR	27-MAY-1999;	99US-0136392.
PR	28-MAY-1999;	99US-0136782.
PR	01-JUN-1999;	99US-0137222.
PR	03-JUN-1999;	99US-0137528.
PR	04-JUN-1999;	99US-0137502.
PR	07-JUN-1999;	99US-0137724.
PR	08-JUN-1999;	99US-0138094.
PR	10-JUN-1999;	99US-0138540.
PR	10-JUN-1999;	99US-0138847.
PR	14-JUN-1999;	99US-0139119.
PR	16-JUN-1999;	99US-0139452.
PR	16-JUN-1999;	99US-0139453.
PR	17-JUN-1999;	99US-0139492.
PR	18-JUN-1999;	99US-0139454.
PR	18-JUN-1999;	99US-0139455.
PR	18-JUN-1999;	99US-0139456.
PR	18-JUN-1999;	99US-0139457.
PR	18-JUN-1999;	99US-0139458.
PR	18-JUN-1999;	99US-0139459.
PR	18-JUN-1999;	99US-0139460.
PR	18-JUN-1999;	99US-0139461.
PR	18-JUN-1999;	99US-0139462.
PR	18-JUN-1999;	99US-0139463.
PR	18-JUN-1999;	99US-0139750.
PR	18-JUN-1999;	99US-0139763.
PR	21-JUN-1999;	99US-0139817.
PR	22-JUN-1999;	99US-0139899.
PR	23-JUN-1999;	99US-0140353.
PR	23-JUN-1999;	99US-0140354.
PR	24-JUN-1999;	99US-0140695.
PR	28-JUN-1999;	99US-0140823.
PR	29-JUN-1999;	99US-0140991.
PR	30-JUN-1999;	99US-0141287.
PR	01-JUL-1999;	99US-0141842.
PR	01-JUL-1999;	99US-0142154.
PR	02-JUL-1999;	99US-0142055.
PR	06-JUL-1999;	99US-0142390.
PR	08-JUL-1999;	99US-0142803.
PR	09-JUL-1999;	99US-0142920.
PR	12-JUL-1999;	99US-0142977.
PR	13-JUL-1999;	99US-0143542.
PR	14-JUL-1999;	99US-0143624.
PR	15-JUL-1999;	99US-0144005.
PR	16-JUL-1999;	99US-0144085.
PR	16-JUL-1999;	99US-0144086.
PR	19-JUL-1999;	99US-0144325.
PR	19-JUL-1999;	99US-0144331.
PR	19-JUL-1999;	99US-0144332.
PR	19-JUL-1999;	99US-0144333.
PR	19-JUL-1999;	99US-0144334.
PR	19-JUL-1999;	99US-0144335.
PR	20-JUL-1999;	99US-0144352.
PR	20-JUL-1999;	99US-0144632.
PR	20-JUL-1999;	99US-0144884.
PR	21-JUL-1999;	99US-0144814.
PR	21-JUL-1999;	99US-0145086.
PR	21-JUL-1999;	99US-0145088.
PR	22-JUL-1999;	99US-0145085.
PR	22-JUL-1999;	99US-0145087.
PR	22-JUL-1999;	99US-0145089.
PR	22-JUL-1999;	99US-0145192.
PR	23-JUL-1999;	99US-0145145.
PR	23-JUL-1999;	99US-0145218.
PR	23-JUL-1999;	99US-0145224.
PR	26-JUL-1999;	99US-0145276.
PR	27-JUL-1999;	99US-0145913.
PR	27-JUL-1999;	99US-0145918.
PR	27-JUL-1999;	99US-0145919.
PR	28-JUL-1999;	99US-0145951.
PR	02-AUG-1999;	99US-0146386.
PR	02-AUG-1999;	99US-0146388.
PR	02-AUG-1999;	99US-0146389.
PR	03-AUG-1999;	99US-0147038.
PR	04-AUG-1999;	99US-0147204.
PR	04-AUG-1999;	99US-0147302.
PR	05-AUG-1999;	99US-0147192.
PR	05-AUG-1999;	99US-0147260.
PR	06-AUG-1999;	99US-0147303.
PR	06-AUG-1999;	99US-0147416.
PR	09-AUG-1999;	99US-0147493.
PR	09-AUG-1999;	99US-0147935.
PR	10-AUG-1999;	99US-0148171.
PR	11-AUG-1999;	99US-0148319.
PR	12-AUG-1999;	99US-0148341.
PR	13-AUG-1999;	99US-0148565.
PR	13-AUG-1999;	99US-0148684.
PR	16-AUG-1999;	99US-0149368.
PR	17-AUG-1999;	99US-0149175.
PR	18-AUG-1999;	99US-0149426.
PR	20-AUG-1999;	99US-0149722.
PR	20-AUG-1999;	99US-0149723.
PR	20-AUG-1999;	99US-0149929.
PR	23-AUG-1999;	99US-0149902.
PR	23-AUG-1999;	99US-0149930.
PR	25-AUG-1999;	99US-0150566.
PR	26-AUG-1999;	99US-0150884.
PR	27-AUG-1999;	99US-0151065.
PR	27-AUG-1999;	99US-0151066.
PR	27-AUG-1999;	99US-0151080.
PR	30-AUG-1999;	99US-0151303.
PR	31-AUG-1999;	99US-0151438.
PR	01-SEP-1999;	99US-0151930.
PR	07-SEP-1999;	99US-0152363.
PR	10-SEP-1999;	99US-0153070.
PR	13-SEP-1999;	99US-0153758.
PR	15-SEP-1999;	99US-0154018.
PR	16-SEP-1999;	99US-0154039.
PR	20-SEP-1999;	99US-0154779.
PR	22-SEP-1999;	99US-0155139.
PR	23-SEP-1999;	99US-0155486.
PR	24-SEP-1999;	99US-0155659.
PR	28-SEP-1999;	99US-0156458.
PR	29-SEP-1999;	99US-0156596.
PR	04-OCT-1999;	99US-0157117.
PR	05-OCT-1999;	99US-0157753.
PR	06-OCT-1999;	99US-0157865.
PR	07-OCT-1999;	99US-0158029.
PR	08-OCT-1999;	99US-0158232.

XX
DE Nucleotide sequence of the second exon of human BMP-16.
XX
KW Human; bone morphogenetic protein-16; BMP-16; murine protein; nodal;
KW formation; bone; cartilage; treatment; wound healing; reduction;
KW fibrosis; scar tissue formation; ss.
XX
OS Homo sapiens.
XX
PN WO9812322-A1.
XX
PD 26-MAR-1998.
XX
PF 09-JUL-1997; 97WO-US11954.
XX
PR 18-SEP-1996; 96US-0715202.
XX
PA (GEMY) GENETICS INST INC.
XX
PI Celeste AJ, Murray BL;
XX
DR WPI; 1998-217262/19.
XX
PT New isolated bone morphogenetic protein-16 - used to develop
PT products for inducing formation of bone, cartilage and other
PT connective tissue, particularly for wound healing and tissue repair
XX
PS Disclosure; Pages 36-37; 43pp; English.
XX
CC The present sequence represents the nucleotide sequence of the second
CC exon of the genomic DNA for human bone morphogenetic protein-16 (BMP-16).
CC Human BMP-16 is a homologue of a murine protein celled nodal, which is
CC expressed in the mouse node during gastrulation. The BMP-16 proteins can
CC induce the formation of bone, cartilage or other connective tissue. They
CC can be used for treating bone, cartilage or other connective tissue
CC defects, periodontal disease or healing of various types of tissues and
CC wounds. They can also increase neuronal, astrocytic and glial cell
CC survival and therefore be useful in transplantation and treatment of
CC conditions exhibiting a decrease in neuronal survival and repair. They
CC can also exhibit properties such as angiogenic, chemotactic and/or
CC chemoattractant properties, and effects on cells including induction of
CC collagen synthesis, fibrosis, differentiation responses, cell
CC proliferative responses and responses involving cell adhesion, migration
CC and extracellular matrices. These properties make the proteins potential
CC agents for wound healing, reduction of fibrosis and reduction of scar
CC tissue formation.
XX
SQ Sequence 2003 BP; 471 A; 490 C; 570 G; 471 T; 1 other;

Query Match 83.2%; Score 15.8; DB 19; Length 2003;
Best Local Similarity 89.5%; Pred. No. 99;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 tttgctcctgccagaggtt 19
| ||||||||| |||||||
Db 802 TCTGCTCCTGCGAGAGGTT 784

RESULT 14
AAT10554
ID AAT10554 standard; cDNA to mRNA; 2818 BP.
XX
AC AAT10554;
XX
DT 03-APR-1996 (first entry)
XX
DE Sheep mammary gland factor cDNA.
XX
KW Mammary gland factor; signal transduction; lactogenic hormone;
KW cytokine regulated transcription factor; transgenic animal; sheep; ss.
XX
OS Ovis aries.
XX
FH Key Location/Qualifiers
FT CDS 244..2628
FT /*tag= a
XX
PN CA2139384-A.
XX
PD 12-NOV-1995.
XX
PF 30-DEC-1994; 94CA-2139384.
XX
PR 11-MAY-1994; 94GB-0009396.
XX
PA (CIBA) CIBA GEIGY AG.
XX
PI Gouilleux F, Groner B, Wakao H;
XX
DR WPI; 1996-077786/09.
DR P-PSDB; AAR88199.
XX
PT DNA encoding mammary gland factor protein - used to identify cpds.
PT affecting intracellular signal transduction of a lactogenic hormone,
PT or cytokine
XX
PS Claim 2; Page 28-34; 42pp; English.
XX
CC A cDNA clone (AAT10554) coding for mammary gland factor (MGF)
CC (AAR88199) was obtd. from a cDNA library derived from sheep
CC lactating mammary tissue mRNA using probes (AAT10555-56) based on
CC internal peptides of MGF. The cDNA is used to produce recombinant
CC MGF in host cells, or as a probe. Transgenic animals, e.g. sheep,
CC overexpressing MGF can be produced in order to increase milk prodn.
CC or to produce a therapeutically useful protein.
XX
SQ Sequence 2818 BP; 618 A; 861 C; 819 G; 520 T; 0 other;

Query Match 83.2%; Score 15.8; DB 17; Length 2818;
Best Local Similarity 89.5%; Pred. No. 1e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 tttgctcctgccagaggtt 19
||| |||||||||||||
Db 2599 tttactcctgccagaggct 2617

RESULT 15
AAC47622
ID AAC47622 standard; DNA; 3346 BP.
XX
AC AAC47622;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 54504.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.

PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 83.2%; Score 15.8; DB 21; Length 3346;
Best Local Similarity 89.5%; Pred. No. 1e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 tttgctcctgccagaggtt 19
||||||||||.|||.||||
Db 1742 tttgctcctggcagtggtt 1760

Search completed: July 25, 2001, 05:23:06
Job time: 4679 sec

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: July 25, 2001, 05:18:53 ; Search time 117.39 Seconds
(without alignments)
29.985 Million cell updates/sec

Title: US-09-142-095-4
Perfect score: 19
Sequence: 1 tttgctcctgccagaggtt 19

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 317530 seqs, 92630169 residues

Total number of hits satisfying chosen parameters: 635060

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
c 1	19	100.0	1190	5	PCT-US92-00282-18	Sequence 18, Appl
c 2	15.8	83.2	843	2	US-08-715-202A-1	Sequence 1, Appli
c 3	15.8	83.2	2002	2	US-08-715-202A-3	Sequence 3, Appli
4	15.8	83.2	2818	1	US-08-366-276-1	Sequence 1, Appli
5	14.8	77.9	2282	1	US-08-373-579-5	Sequence 5, Appli
6	14.8	77.9	2282	1	US-08-418-595-5	Sequence 5, Appli
7	14.8	77.9	2282	2	US-08-665-926-5	Sequence 5, Appli
8	14.8	77.9	2282	4	US-09-162-437-5	Sequence 5, Appli
9	14.8	77.9	5924	1	US-08-447-411-44	Sequence 44, Appl
10	14.8	77.9	5948	2	US-08-662-227-1	Sequence 1, Appli
c 11	14.4	75.8	1858	2	US-08-359-705B-7	Sequence 7, Appli
c 12	14.4	75.8	1858	2	US-08-286-846A-7	Sequence 7, Appli
c 13	14.4	75.8	1858	2	US-08-457-880A-7	Sequence 7, Appli
c 14	14.4	75.8	1858	3	US-08-444-622A-7	Sequence 7, Appli
c 15	14.4	75.8	1858	3	US-08-942-562-7	Sequence 7, Appli
c 16	14.4	75.8	1858	4	US-09-156-923-7	Sequence 7, Appli
c 17	14.4	75.8	2715	2	US-08-359-705B-5	Sequence 5, Appli
c 18	14.4	75.8	2715	2	US-08-286-846A-5	Sequence 5, Appli
c 19	14.4	75.8	2715	2	US-08-457-880A-5	Sequence 5, Appli
c 20	14.4	75.8	2715	3	US-08-444-622A-5	Sequence 5, Appli
c 21	14.4	75.8	2715	3	US-08-942-562-5	Sequence 5, Appli
c 22	14.4	75.8	2715	4	US-09-156-923-5	Sequence 5, Appli
c 23	14.4	75.8	2940	1	US-08-286-305A-8	Sequence 8, Appli
c 24	14.4	75.8	2940	2	US-08-441-104A-8	Sequence 8, Appli
c 25	14.4	75.8	2940	2	US-08-440-816A-8	Sequence 8, Appli
c 26	14.2	74.7	300	2	US-07-938-154-6	Sequence 6, Appli
c 27	14.2	74.7	300	5	PCT-US91-02311-6	Sequence 6, Appli
28	14.2	74.7	693	1	US-08-052-205-10	Sequence 10, Appl
29	14.2	74.7	693	1	US-08-595-974-10	Sequence 10, Appl
30	14.2	74.7	759	1	US-08-052-205-8	Sequence 8, Appli
31	14.2	74.7	759	1	US-08-595-974-8	Sequence 8, Appli
32	14.2	74.7	1044	1	US-08-052-205-6	Sequence 6, Appli
33	14.2	74.7	1044	1	US-08-595-974-6	Sequence 6, Appli
34	14.2	74.7	1062	1	US-08-052-205-1	Sequence 1, Appli
35	14.2	74.7	1062	1	US-08-595-974-1	Sequence 1, Appli
36	14.2	74.7	1110	1	US-08-052-205-5	Sequence 5, Appli
37	14.2	74.7	1110	1	US-08-595-974-5	Sequence 5, Appli
38	14.2	74.7	1393	1	US-08-052-205-2	Sequence 2, Appli
39	14.2	74.7	1393	1	US-08-595-974-2	Sequence 2, Appli
40	14.2	74.7	1449	2	US-08-705-868-2	Sequence 2, Appli
41	14.2	74.7	1449	3	US-09-123-615-2	Sequence 2, Appli
42	14.2	74.7	1451	1	US-08-031-143B-68	Sequence 68, Appl
43	14.2	74.7	1470	1	US-08-052-205-3	Sequence 3, Appli
44	14.2	74.7	1470	1	US-08-595-974-3	Sequence 3, Appli
45	14.2	74.7	1608	2	US-08-424-224-1	Sequence 1, Appli

ALIGNMENTS

```
RESULT 1
PCT-US92-00282-18/c
; Sequence 18, Application PC/TUS9200282
;  GENERAL INFORMATION:
;    APPLICANT:  OWENS, IDA S.
;    APPLICANT:  RITTER, JOSEPH K.
;    TITLE OF INVENTION:  THE GENETIC LOCUS UGT1 AND A MUTATION
;    TITLE OF INVENTION:  THEREIN.
;    NUMBER OF SEQUENCES:  40
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN DARBY & CUSHMAN
;      STREET:  1615 L STREET, N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20036-5601
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US92/00282
;      FILING DATE:  19920110
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SCOTT, WATSON T.
;      REGISTRATION NUMBER:  26581
;      REFERENCE/DOCKET NUMBER:  91532-PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-861-3000
;      TELEFAX:  202-822-0944
;      TELEX:  6714627 CUSH
;  INFORMATION FOR SEQ ID NO:  18:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1190 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
PCT-US92-00282-18


  Query Match           100.0%;  Score 19;  DB 5;  Length 1190;
  Best Local Similarity  100.0%;  Pred. No. 0.49;
  Matches   19;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 tttgctcctgccagaggtt 19
           |||||||||||||||||||
Db      78 TTTGCTCCTGCCAGAGGTT 60
```

```
RESULT   2
US-08-715-202A-1/c
; Sequence 1, Application US/08715202A
; Patent No. 5965403
;  GENERAL INFORMATION:
;    APPLICANT:  CELESTE, ANTHONY J.
;    APPLICANT:  MURRAY, BETH L.
;    TITLE OF INVENTION:  BONE MORPHOGENETIC PROTEIN-16 (BMP-16)
;    TITLE OF INVENTION:  COMPOSITIONS
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  GENETICS INSTITUTE, INC.
;      STREET:  87 CAMBRIDGEPARK DRIVE
;      CITY:  CAMBRIDGE
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02140
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/715,202A
;      FILING DATE:  September 18, 1996
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  LAZAR, STEVEN R
;      REGISTRATION NUMBER:  32,618
;      REFERENCE/DOCKET NUMBER:  5275
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 498-8260
;      TELEFAX:  (617) 876-5851
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  843 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  sig_peptide
;      LOCATION:  1..510
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  511..840
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..840
US-08-715-202A-1


  Query Match             83.2%;  Score 15.8;  DB 2;  Length 843;
  Best Local Similarity   89.5%;  Pred. No. 20;
  Matches   17;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 tttgctcctgccagaggtt 19
           | ||||||||| |||||||
Db     412 TCTGCTCCTGCGAGAGGTT 394


RESULT   3
US-08-715-202A-3/c
; Sequence 3, Application US/08715202A
; Patent No. 5965403
;  GENERAL INFORMATION:
;    APPLICANT:  CELESTE, ANTHONY J.
;    APPLICANT:  MURRAY, BETH L.
;    TITLE OF INVENTION:  BONE MORPHOGENETIC PROTEIN-16 (BMP-16)
;    TITLE OF INVENTION:  COMPOSITIONS
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  GENETICS INSTITUTE, INC.
;      STREET:  87 CAMBRIDGEPARK DRIVE
;      CITY:  CAMBRIDGE
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02140
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/715,202A
;      FILING DATE:  September 18, 1996
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  LAZAR, STEVEN R
;      REGISTRATION NUMBER:  32,618
;      REFERENCE/DOCKET NUMBER:  5275
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 498-8260
;      TELEFAX:  (617) 876-5851
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2002 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-715-202A-3


  Query Match             83.2%;  Score 15.8;  DB 2;  Length 2002;
  Best Local Similarity   89.5%;  Pred. No. 23;
  Matches   17;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 tttgctcctgccagaggtt 19
           | ||||||||| |||||||
Db     801 TCTGCTCCTGCGAGAGGTT 783


RESULT   4
US-08-366-276-1
; Sequence 1, Application US/08366276
; Patent No. 5534409
;  GENERAL INFORMATION:
;    APPLICANT:  Groner, Bernd
;    APPLICANT:  Gouilleux, Fabrice
;    APPLICANT:  Wakao, Hiroshi
;    TITLE OF INVENTION:  Cytokine Regulated Transcription Factor
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CIBA-GEIGY Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30B
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/366,276
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9409396.0
;      FILING DATE:  11-MAY-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Elmer, James Scoot
```

```
;      REGISTRATION NUMBER:  36,129
;      REFERENCE/DOCKET NUMBER:  4-19992/A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (919) 541-8614
;      TELEFAX:  (919) 541-8689
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2818 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  244..2625
;      OTHER INFORMATION:  /product= "mammary gland factor"
US-08-366-276-1


  Query Match             83.2%;  Score 15.8;  DB 1;  Length 2818;
  Best Local Similarity   89.5%;  Pred. No. 24;
  Matches   17;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 tttgctcctgccagaggtt 19
           ||| |||||||||||||  |
Db    2599 TTTACTCCTGCCAGAGGCT 2617


RESULT   5
US-08-373-579-5
; Sequence 5, Application US/08373579
; Patent No. 5650490
;  GENERAL INFORMATION:
;    APPLICANT:  Davis, et al.
;    TITLE OF INVENTION:  TIE-2 LIGAND, METHOD OF MAKING AND USES
;    TITLE OF INVENTION:  THEREOF
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Regeneron Pharmaceuticals, Inc.
;      STREET:  777 Old Saw Mill River Road
;      CITY:  Tarrytown
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10591
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/373,579
;      FILING DATE:  17-JAN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/353,503
;      FILING DATE:  09-DEC-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/348,492
;      FILING DATE:  02-DEC-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/330,261
;      FILING DATE:  27-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/319,932
;      FILING DATE:  07-OCT-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Cobert, Robert J.
;      REGISTRATION NUMBER:  36,108
;      REFERENCE/DOCKET NUMBER:  REG 330-D
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (914) 345-7400
;      TELEFAX:  (914) 345-7721
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2282 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  357..1847
US-08-373-579-5


  Query Match             77.9%;  Score 14.8;  DB 1;  Length 2282;
  Best Local Similarity   88.9%;  Pred. No. 75;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 tttgctcctgccagaggt 18
           ||| |||||||||||| |
Db    488 TTTCCTCCTGCCAGAGAT 505


RESULT   6
US-08-418-595-5
; Sequence 5, Application US/08418595
; Patent No. 5814464
;  GENERAL INFORMATION:
;    APPLICANT:  Davis, et al.
;    TITLE OF INVENTION:  TIE-2 LIGAND, METHOD OF MAKING AND USES
;    TITLE OF INVENTION:  THEREOF
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Regeneron Pharmaceuticals, Inc.
;      STREET:  777 Old Saw Mill River Road
;      CITY:  Tarrytown
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10591
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/418,595
;      FILING DATE:  06-APR-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/373,579
;      FILING DATE:  17-JAN-1995
;      APPLICATION NUMBER:  US 08/353,503
;      FILING DATE:  09-DEC-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/348,492
;      FILING DATE:  02-DEC-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/330,261
;      FILING DATE:  27-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/319,932
;      FILING DATE:  07-OCT-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Cobert, Robert J.
;      REGISTRATION NUMBER:  36,108
;      REFERENCE/DOCKET NUMBER:  REG 330-D
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (914) 345-7400
;      TELEFAX:  (914) 345-7721
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2282 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
```

```
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  357..1847
US-08-418-595-5


  Query Match             77.9%;  Score 14.8;  DB 1;  Length 2282;
  Best Local Similarity   88.9%;  Pred. No. 75;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 tttgctcctgccagaggt 18
          ||| |||||||||||| |
Db    488 TTTCCTCCTGCCAGAGAT 505


RESULT   7
US-08-665-926-5
; Sequence 5, Application US/08665926
; Patent No. 5851797
;  GENERAL INFORMATION:
;    APPLICANT:  Valenzuela et al.
;    TITLE OF INVENTION:  TIE LIGAND-3, METHODS OF MAKING AND USES
;    TITLE OF INVENTION:  THEREOF
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Regeneron Pharmaceuticals, Inc.
;      STREET:  777 Old Saw Mill River Road
;      CITY:  Tarrytown
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10591-6707
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/665,926
;      FILING DATE:  19-JUN-1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Robert J. Cobert
;      REGISTRATION NUMBER:  36,108
;      REFERENCE/DOCKET NUMBER:  REG 330-H
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (914) 345-7400
;      TELEFAX:  (914) 345-2113
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2282 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  357..1847
US-08-665-926-5


  Query Match             77.9%;  Score 14.8;  DB 2;  Length 2282;
  Best Local Similarity   88.9%;  Pred. No. 75;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 tttgctcctgccagaggt 18
          ||| |||||||||||| |
Db    488 TTTCCTCCTGCCAGAGAT 505


RESULT   8
US-09-162-437-5
; Sequence 5, Application US/09162437
; Patent No. 6166185
;  GENERAL INFORMATION:
;    APPLICANT:  Davis, et al.
;    TITLE OF INVENTION:  TIE-2 LIGAND, METHOD OF MAKING AND USES
;    TITLE OF INVENTION:  THEREOF
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Regeneron Pharmaceuticals, Inc.
;      STREET:  777 Old Saw Mill River Road
;      CITY:  Tarrytown
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10591
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/162,437
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/418,595
;      FILING DATE:  06-APR-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/373,579
;      FILING DATE:  17-JAN-1995
;      APPLICATION NUMBER:  US 08/353,503
;      FILING DATE:  09-DEC-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/348,492
;      FILING DATE:  02-DEC-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/330,261
;      FILING DATE:  27-OCT-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/319,932
;      FILING DATE:  07-OCT-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Cobert, Robert J.
;      REGISTRATION NUMBER:  36,108
;      REFERENCE/DOCKET NUMBER:  REG 330-D
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (914) 345-7400
;      TELEFAX:  (914) 345-7721
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2282 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  357..1847
US-09-162-437-5


  Query Match             77.9%;  Score 14.8;  DB 4;  Length 2282;
  Best Local Similarity   88.9%;  Pred. No. 75;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 tttgctcctgccagaggt 18
          ||| |||||||||||| |
Db    488 TTTCCTCCTGCCAGAGAT 505


RESULT   9
US-08-447-411-44
; Sequence 44, Application US/08447411
; Patent No. 5773243
```

```
;  GENERAL INFORMATION:
;    APPLICANT:  FRITZINGER, DAVID C.
;    APPLICANT:  BREDEHORST, REINHARD
;    APPLICANT:  VOGEL, CARL-WILHELM
;    TITLE OF INVENTION:  DNA ENCODING COBRA C3, CVF1, AND CVF2
;    NUMBER OF SEQUENCES:  81
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 S. Jefferson Davis Highway, Suite 400
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/447,411
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/043,747
;      FILING DATE:  07-APR-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Oblon, No. 5773243man F.
;      REGISTRATION NUMBER:  24,618
;      REFERENCE/DOCKET NUMBER:  1126-101-0
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 413-3000
;      TELEFAX:  (703) 413-2220
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  44:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5924 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  sig_peptide
;      LOCATION:  4..69
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  70..4929
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  4..4929
US-08-447-411-44


  Query Match           77.9%;  Score 14.8;  DB 1;  Length 5924;
  Best Local Similarity   88.9%;  Pred. No. 87;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 tttgctcctgccagaggt 18
           || ||||||| |||||||
Db    5567 TTGGCTCCTGACAGAGGT 5584


RESULT  10
US-08-662-227-1
; Sequence 1, Application US/08662227
; Patent No. 5922320
;  GENERAL INFORMATION:
;    APPLICANT:  VOGEL, CARL-WILHELM
;    APPLICANT:  BREDEHORST, REINHORST
;    APPLICANT:  KOCK, MICHAEL
;    APPLICANT:  FRITZINGER, DAVID
;    TITLE OF INVENTION:  RECOMBINANT PROCVF
;    NUMBER OF SEQUENCES:  39
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 S. JEFFERSON DAVIS HIGHWAY
;      CITY:  ARLINGTON
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/662,227
;      FILING DATE:  14-JUN-1996
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  OBLON, NORMAN F.
;      REGISTRATION NUMBER:  24,618
;      REFERENCE/DOCKET NUMBER:  1126-0107-0X
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-413-3000
;      TELEFAX:  703-413-2220
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5948 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-662-227-1


  Query Match           77.9%;  Score 14.8;  DB 2;  Length 5948;
  Best Local Similarity   88.9%;  Pred. No. 87;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 tttgctcctgccagaggt 18
           || ||||||| |||||||
Db    5567 TTGGCTCCTGACAGAGGT 5584


RESULT  11
US-08-359-705B-7/c
; Sequence 7, Application US/08359705B
; Patent No. 5844092
;  GENERAL INFORMATION:
;    APPLICANT:  Presta, Leonard G.
;    APPLICANT:  Shelton, David L.
;    APPLICANT:  Urfer, Roman
;    TITLE OF INVENTION:  Human trk Receptors and Neurotrophic Factor Inhibitors
;    NUMBER OF SEQUENCES:  41
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  1 DNA Way
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 inch, 1.44 Mb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WinPatin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/359,705B
;      FILING DATE:  20-Dec-1994
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/286846
;      FILING DATE:  08/10/94
;    PRIOR APPLICATION DATA:
```

```
;       APPLICATION NUMBER:  08/215139
;       FILING DATE:  03/18/94
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Torchia, PhD., Timothy E.
;       REGISTRATION NUMBER:  36,700
;       REFERENCE/DOCKET NUMBER:  P0873P2
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  650/225-8674
;       TELEFAX:  650/952-9881
;   INFORMATION FOR SEQ ID NO:  7:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1858 base pairs
;       TYPE:  Nucleic Acid
;       STRANDEDNESS:  Single
;       TOPOLOGY:  Linear
US-08-359-705B-7


  Query Match           75.8%;  Score 14.4;  DB 2;  Length 1858;
  Best Local Similarity  93.8%;  Pred. No. 1.2e+02;
  Matches   15;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      3 tgctcctgccagaggt 18
          ||||||||||||||| |
Db    545 TGCTCCTGCCAGAGCT 530


RESULT  12
US-08-286-846A-7/c
; Sequence 7, Application US/08286846A
; Patent No. 5877016
;  GENERAL INFORMATION:
;    APPLICANT:  Presta, Leonard G.
;    APPLICANT:  Shelton, David L.
;    APPLICANT:  Urfer, Roman
;    TITLE OF INVENTION:  Human trk Receptors and Neurotrophic Factor Inhibitors
;    NUMBER OF SEQUENCES:  41
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 inch, 1.44 Mb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WinPatin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/286,846A
;      FILING DATE:  05-Aug-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Torchia, PhD., Timothy E.
;      REGISTRATION NUMBER:  36,700
;      REFERENCE/DOCKET NUMBER:  P0873P1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/225-8674
;      TELEFAX:  415/952-9881
;      TELEX:  910/371-7168
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1858 base pairs
;      TYPE:  Nucleic Acid
;      STRANDEDNESS:  Single
;      TOPOLOGY:  Linear
US-08-286-846A-7


  Query Match           75.8%;  Score 14.4;  DB 2;  Length 1858;
  Best Local Similarity  93.8%;  Pred. No. 1.2e+02;
  Matches   15;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      3 tgctcctgccagaggt 18
          ||||||||||||||| |
Db    545 TGCTCCTGCCAGAGCT 530


RESULT  13
US-08-457-880A-7/c
; Sequence 7, Application US/08457880A
; Patent No. 5910574
;  GENERAL INFORMATION:
;    APPLICANT:  Leonard G. Presta
;    APPLICANT:  David L. Shelton
;    APPLICANT:  Roman Urfer
;    TITLE OF INVENTION:  HUMAN trk RECEPTORS AND NEUROTROPHIC FACTOR
;    TITLE OF INVENTION:  INHIBITORS
;    NUMBER OF SEQUENCES:  41
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  1 DNA Way
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 inch, 1.44 Mb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WinPatin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/457,880A
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/444,622
;      FILING DATE:  19-May-1995
;      APPLICATION NUMBER:  08/286846
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Torchia, PhD., Timothy E.
;      REGISTRATION NUMBER:  36,700
;      REFERENCE/DOCKET NUMBER:  P0873P1C3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650/225-8674
;      TELEFAX:  650/952-9881
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1858 base pairs
;      TYPE:  Nucleic Acid
;      STRANDEDNESS:  Single
;      TOPOLOGY:  Linear
US-08-457-880A-7


  Query Match           75.8%;  Score 14.4;  DB 2;  Length 1858;
  Best Local Similarity  93.8%;  Pred. No. 1.2e+02;
  Matches   15;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      3 tgctcctgccagaggt 18
          ||||||||||||||| |
Db    545 TGCTCCTGCCAGAGCT 530


RESULT  14
US-08-444-622A-7/c
; Sequence 7, Application US/08444622A
; Patent No. 6025166
;  GENERAL INFORMATION:
;    APPLICANT:  Leonard G. Presta
;    APPLICANT:  David L. Shelton
;    APPLICANT:  Roman Urfer
;    TITLE OF INVENTION:  HUMAN trk RECEPTORS AND NEUROTROPHIC FACTOR INHIBITORS
;    NUMBER OF SEQUENCES:  41
```

```
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  1 DNA Way
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 inch, 1.44 Mb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WinPatin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/444,622A
;      FILING DATE:  19-May-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/286846
;      FILING DATE:  5
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Torchia, PhD., Timothy E.
;      REGISTRATION NUMBER:  36,700
;      REFERENCE/DOCKET NUMBER:  P0873P1C3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650/225-8674
;      TELEFAX:  650/952-9881
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1858 base pairs
;      TYPE:  Nucleic Acid
;      STRANDEDNESS:  Single
;      TOPOLOGY:  Linear
US-08-444-622A-7


  Query Match           75.8%;  Score 14.4;  DB 3;  Length 1858;
  Best Local Similarity  93.8%;  Pred. No. 1.2e+02;
  Matches   15;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      3 tgctcctgccagaggt 18
          ||||||||||||||| |
Db    545 TGCTCCTGCCAGAGCT 530


RESULT  15
US-08-942-562-7/c
; Sequence 7, Application US/08942562
; Patent No. 6027927
;  GENERAL INFORMATION:
;    APPLICANT:  Presta, Leonard G.
;    APPLICANT:  Shelton, David L.
;    APPLICANT:  Urfer, Roman
;    TITLE OF INVENTION:  Human trk Receptors and Neurotrophic
;    TITLE OF INVENTION:  Factor Inhibitors
;    NUMBER OF SEQUENCES:  37
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 inch, 1.44 Mb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WinPatin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/942,562
;      FILING DATE:  01-OCT-1997
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/444,597
;      FILING DATE:  19-May-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Torchia, PhD., Timothy E.
;      REGISTRATION NUMBER:  36,700
;      REFERENCE/DOCKET NUMBER:  P0873P1C2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/225-8674
;      TELEFAX:  415/952-9881
;      TELEX:  910/371-7168
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1858 base pairs
;      TYPE:  Nucleic Acid
;      STRANDEDNESS:  Single
;      TOPOLOGY:  Linear
US-08-942-562-7


  Query Match           75.8%;  Score 14.4;  DB 3;  Length 1858;
  Best Local Similarity  93.8%;  Pred. No. 1.2e+02;
  Matches   15;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      3 tgctcctgccagaggt 18
          ||||||||||||||| |
Db    545 TGCTCCTGCCAGAGCT 530


Search completed: July 25, 2001, 05:18:55
Job time: 9218 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: July 25, 2001, 02:01:07 ; Search time 2762.24 Seconds
(without alignments)
71.866 Million cell updates/sec

Title: US-09-142-095-1
Perfect score: 21
Sequence: 1 aagtgaactccctgctacctt 21

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est33:*
26: gb_est34:*
27: gb_est35:*
28: gb_est36:*
29: gb_est37:*
30: gb_est38:*
31: gb_est39:*
32: gb_est40:*
33: em_estba:*
34: em_estfun:*
35: em_esthum1:*
36: em_esthum2:*
37: em_esthum3:*
38: em_esthum4:*
39: em_esthum5:*
40: em_esthum6:*
41: em_esthum7:*
42: em_esthum8:*
43: em_esthum9:*
44: em_esthum10:*
45: em_esthum11:*
46: em_esthum12:*
47: em_esthum13:*
48: em_esthum14:*
49: em_esthum15:*
50: em_esthum16:*
51: em_esthum17:*
52: em_esthum18:*
53: em_esthum19:*
54: em_esthum20:*
55: em_esthum21:*
56: em_esthum22:*
57: em_esthum23:*
58: em_esthum24:*
59: em_esthum25:*
60: em_esthum26:*
61: em_esthum27:*
62: em_esthum28:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estin5:*
68: em_estom1:*
69: em_estom2:*
70: em_estov1:*
71: em_estov2:*
72: em_estpl1:*
73: em_estpl2:*
74: em_estpl3:*
75: em_estpl4:*
76: em_estpl5:*
77: em_estpl6:*
78: em_estpl7:*
79: em_estpl8:*
80: em_estpl9:*
81: em_estpl10:*
82: em_estro1:*
83: em_estro2:*
84: em_estro3:*
85: em_estro4:*
86: em_estro5:*
87: em_estro6:*
88: em_estro7:*
89: em_estro8:*
90: em_estro9:*
91: em_estro10:*
92: em_estro11:*
93: em_estro12:*
94: em_estro13:*
95: em_estro14:*
96: em_estro15:*
97: em_estro16:*
98: em_estro17:*
99: em_estro18:*
100: em_estro19:*
101: em_estro20:*
102: gb_est25:*
103: gb_est26:*
104: gb_est27:*
105: gb_est28:*
106: gb_est29:*
107: gb_est30:*
108: gb_est31:*
109: gb_est32:*
110: gb_est41:*
111: gb_est42:*
112: gb_est43:*
113: gb_est44:*
114: gb_est45:*
115: gb_est46:*
116: gb_est47:*

gb_est48:*
gb_est49:*
gb_est50:*
gb_est51:*
gb_est52:*
gb_est53:*
gb_est54:*
gb_est55:*
gb_est56:*
gb_est57:*
gb_est58:*
gb_est59:*
gb_est60:*
gb_est61:*
gb_est62:*
gb_est63:*
gb_est64:*
gb_est65:*
gb_est66:*
gb_est75:*
gb_est76:*
gb_est77:*
gb_est78:*
gb_est79:*
gb_est80:*
gb_est81:*
gb_est82:*
gb_est83:*
gb_est84:*
gb_est85:*
gb_est86:*
gb_est87:*
gb_est88:*
gb_est89:*
gb_est90:*
gb_est99:*
gb_est100:*
gb_est101:*
gb_est102:*
gb_est103:*
gb_est104:*
gb_est105:*
gb_est106:*
gb_est67:*
gb_est68:*
gb_est69:*
gb_est70:*
gb_est71:*
gb_est72:*
gb_est73:*
gb_est74:*
gb_est91:*
gb_est92:*
gb_est93:*
gb_est94:*
gb_est95:*
gb_est96:*
gb_est97:*
gb_est98:*
em_esthum29:*
em_esthum30:*
em_esthum31:*
em_esthum32:*
em_esthum33:*
em_estom3:*
em_estpl11:*
em_estro21:*
em_estro22:*
em_estro23:*
em_htc:*
gb_est107:*
gb_est108:*
gb_est109:*
gb_est110:*
gb_est111:*
gb_htc:*
em_gss_fun:*
em_gss_hum1:*
em_gss_hum2:*
em_gss_hum3:*
em_gss_hum4:*
em_gss_hum5:*
em_gss_hum6:*
em_gss_hum7:*
em_gss_hum8:*
em_gss_hum9:*
em_gss_inv1:*
em_gss_inv2:*
em_gss_inv3:*
em_gss_other:*
em_gss_pln1:*
em_gss_pln2:*
em_gss_pro:*
em_gss_rod1:*
em_gss_rod2:*
em_gss_rod3:*
em_gss_rod4:*
em_gss_rod5:*
em_gss_vrt1:*
em_gss_vrt2:*
em_gss_vrt3:*
gb_gss1:*
gb_gss2:*
gb_gss3:*
gb_gss4:*
gb_gss5:*
gb_gss6:*
gb_gss7:*
gb_gss8:*
gb_gss9:*
gb_gss10:*
gb_gss11:*
gb_gss12:*
gb_gss13:*
gb_gss14:*
gb_gss15:*
gb_gss16:*
gb_gss17:*
gb_gss18:*
gb_gss19:*
gb_gss20:*
gb_gss21:*
gb_gss22:*
gb_gss23:*
gb_gss24:*
gb_gss25:*
gb_gss26:*
gb_gss27:*
gb_gss28:*
gb_gss29:*
gb_gss30:*
gb_gss31:*
gb_gss32:*
gb_gss33:*
gb_gss34:*
em_gss_inv4:*
em_gss_rod6:*
em_gss_rod7:*
em_gss_rod8:*
gb_gss35:*
gb_gss36:*
gb_gss37:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	17.8	84.8	212	188	R68057	R68057 yh98e12.r1
2	17.8	84.8	348	187	R31201	R31201 yh62a02.r1
3	17.8	84.8	417	188	R67982	R67982 yi04a10.r1
4	17.8	84.8	418	187	R31714	R31714 yh63a12.r1
5	17.8	84.8	481	188	R63154	R63154 yi01a11.r1
6	17.8	84.8	498	235	AQ882343	AQ882343 HS_5382_A
c 7	17.8	84.8	510	225	AQ185931	AQ185931 HS_3065_A
8	17.8	84.8	608	230	AQ551067	AQ551067 RPCI-11-4
c 9	17	81.0	471	31	AV553454	AV553454 AV553454
c 10	16.8	80.0	193	115	AW381670	AW381670 QV0-HT031
11	16.8	80.0	304	8	AA500281	AA500281 vi97f08.r
12	16.8	80.0	326	9	AA560193	AA560193 vl20g01.r
13	16.8	80.0	410	9	AA620205	AA620205 vo64a06.r
14	16.8	80.0	416	22	AI615361	AI615361 vi97f08.y
15	16.8	80.0	439	9	AA616429	AA616429 vo09e02.r
16	16.8	80.0	469	8	AA473687	AA473687 vg89c05.r
17	16.8	80.0	508	13	AA881891	AA881891 vx29b10.r
c 18	16.8	80.0	514	236	AQ978201	AQ978201 RPCI-23-2
c 19	16.8	80.0	535	238	AZ093040	AZ093040 RPCI-23-4
20	16.8	80.0	547	120	AW743288	AW743288 up64h01.y
21	16.8	80.0	556	2	AA108719	AA108719 mp29f06.r
c 22	16.8	80.0	561	238	AZ123144	AZ123144 RPCI-23-4
c 23	16.8	80.0	740	245	AZ561670	AZ561670 RPCI-23-1
c 24	16.4	78.1	322	251	AZ902873	AZ902873 RPCI-24-1
c 25	16.4	78.1	415	147	BF370208	BF370208 RC4-FN003
c 26	16.4	78.1	481	114	AW342557	AW342557 GthEST414
c 27	16.4	78.1	508	21	AI499582	AI499582 tn96e01.x
c 28	16.4	78.1	530	123	AW981762	AW981762 PC18D09 P
c 29	16.4	78.1	665	240	AZ263471	AZ263471 RPCI-23-1
c 30	16.4	78.1	668	246	AZ617251	AZ617251 1M0448M21
31	16.4	78.1	783	155	BG573720	BG573720 602594785
c 32	16.4	78.1	810	221	CNS03GRY	AL243367 Tetraodon
c 33	16.4	78.1	911	147	BF304665	BF304665 601888071
c 34	16.4	78.1	914	220	CNS01TC5	AL166334 Tetraodon
c 35	16.4	78.1	1049	222	CNS05R8X	AL350250 Tetraodon
36	16.2	77.1	193	218	AF149685	AF149685 AF149685
c 37	16.2	77.1	265	11	AA716271	AA716271 zh28d11.s
38	16.2	77.1	274	23	AI651623	AI651623 wb07d07.x
39	16.2	77.1	285	15	AI021546	AI021546 ub09c11.r
c 40	16.2	77.1	319	14	AA985674	AA985674 or71e07.s
41	16.2	77.1	340	24	AI744438	AI744438 wf89c03.x
c 42	16.2	77.1	351	4	AA235220	AA235220 zs36e12.r
43	16.2	77.1	355	123	AW999271	AW999271 MR0-BN007
44	16.2	77.1	370	18	AI274481	AI274481 qv61d06.x
45	16.2	77.1	378	168	BF722600	BF722600 mab21c09.

ALIGNMENTS

RESULT 1
R68057
LOCUS R68057 212 bp mRNA EST 01-JUN-1995
DEFINITION yh98e12.r1 Soares placenta Nb2HP Homo sapiens cDNA clone IMAGE:137806 5', mRNA sequence.
ACCESSION R68057
VERSION R68057.1 GI:841574
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 212)
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and Wilson,R.
TITLE The WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Insert Size: 761
High quality sequence stops: 74
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 761 Std Error: 0.00
Seq primer: M13RP1
High quality sequence stop: 74.
FEATURES Location/Qualifiers
source 1. .212
/organism="Homo sapiens"
/db_xref="GDB:544134"
/db_xref="taxon:9606"
/clone="IMAGE:137806"
/clone_lib="Soares placenta Nb2HP"
/sex="Female"
/dev_stage="placenta obtained at birth (full term)"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: placenta; Vector: pT7T3D (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' AACTGGAAGAATTCGCGGCCGCAGGAATTTTTTTTTTTTTTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo. "
BASE COUNT 50 a 63 c 32 g 66 t 1 others
ORIGIN

Query Match 84.8%; Score 17.8; DB 188; Length 212;
Best Local Similarity 90.5%; Pred. No. 1.4e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy       1 aagtgaactccctgctacctt 21
           |||| | ||||||||||||||
Db      11 AAGTTAGCTCCCTGCTACCTT 31
```

RESULT 2
R31201
LOCUS R31201 348 bp mRNA EST 28-APR-1995
DEFINITION yh62a02.r1 Soares placenta Nb2HP Homo sapiens cDNA clone IMAGE:134282 5' similar to contains MSR1 repetitive element ;, mRNA sequence.
ACCESSION R31201
VERSION R31201.1 GI:787044
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 348)
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and Wilson,R.
TITLE The WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810

Email: est@watson.wustl.edu
Insert Size: 728
High quality sequence stops: 195
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 728 Std Error: 0.00
Seq primer: M13RP1
High quality sequence stop: 195.

```
FEATURES             Location/Qualifiers
     source          1. .348
                     /organism="Homo sapiens"
                     /db_xref="GDB:540090"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:134282"
                     /clone_lib="Soares placenta Nb2HP"
                     /sex="Female"
                     /dev_stage="placenta obtained at birth (full term)"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Organ: placenta; Vector: pT7T3D (Pharmacia) with a
                     modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer [5'
                     AACTGGAAGAATTCGCGGCCGCAGGAATTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization. Library
                     constructed by Bento Soares and M.Fatima Bonaldo. "
BASE COUNT       78 a     92 c     66 g    108 t      4 others
ORIGIN


  Query Match             84.8%;  Score 17.8;  DB 187;  Length 348;
  Best Local Similarity   90.5%;  Pred. No. 1.5e+02;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 aagtgaactccctgctacctt 21
           |||| | |||||||||||||||
Db      10 AAGTTAGCTCCCTGCTACCTT 30


RESULT   3
R67982
LOCUS       R67982         417 bp    mRNA            EST       01-JUN-1995
DEFINITION  yi04a10.r1 Soares placenta Nb2HP Homo sapiens cDNA clone
            IMAGE:138234 5', mRNA sequence.
ACCESSION   R67982
VERSION     R67982.1  GI:841499
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 417)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman
            ,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J.,
            Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston
            ,R., Williamson,A., Wohldmann,P. and Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Insert Size: 774
            High quality sequence stops: 306
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 774  Std Error: 0.00
            Seq primer: M13RP1
            High quality sequence stop: 306.
FEATURES             Location/Qualifiers
     source          1. .417
                     /organism="Homo sapiens"
                     /db_xref="GDB:544609"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:138234"
                     /clone_lib="Soares placenta Nb2HP"
                     /sex="Female"
                     /dev_stage="placenta obtained at birth (full term)"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Organ: placenta; Vector: pT7T3D (Pharmacia) with a
                     modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer [5'
                     AACTGGAAGAATTCGCGGCCGCAGGAATTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization. Library
                     constructed by Bento Soares and M.Fatima Bonaldo. "
BASE COUNT      104 a    101 c     89 g    123 t
ORIGIN


  Query Match             84.8%;  Score 17.8;  DB 188;  Length 417;
  Best Local Similarity   90.5%;  Pred. No. 1.6e+02;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 aagtgaactccctgctacctt 21
           |||| | |||||||||||||||
Db      10 AAGTTAGCTCCCTGCTACCTT 30


RESULT   4
R31714
LOCUS       R31714         418 bp    mRNA            EST       28-APR-1995
DEFINITION  yh63a12.r1 Soares placenta Nb2HP Homo sapiens cDNA clone
            IMAGE:134398 5', mRNA sequence.
ACCESSION   R31714
VERSION     R31714.1  GI:787557
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 418)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman
            ,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J.,
            Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston
            ,R., Williamson,A., Wohldmann,P. and Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Insert Size: 707
            High quality sequence stops: 244
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 707  Std Error: 0.00
            Seq primer: M13RP1
            High quality sequence stop: 244.
FEATURES             Location/Qualifiers
     source          1. .418
                     /organism="Homo sapiens"
                     /db_xref="GDB:540219"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:134398"
                     /clone_lib="Soares placenta Nb2HP"
```

/sex="Female"
/dev_stage="placenta obtained at birth (full term)"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: placenta; Vector: pT7T3D (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' AACTGGAAGAATTCGCGGCCGCAGGAATTTTTTTTTTTTTTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo. "
BASE COUNT 90 a 100 c 95 g 126 t 7 others
ORIGIN

Query Match 84.8%; Score 17.8; DB 187; Length 418;
Best Local Similarity 90.5%; Pred. No. 1.6e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 aagtgaactccctgctacctt 21
|||| | ||||||||||||||||
Db 10 AAGTTAGCTCCCTGCTACCTT 30

RESULT 5
R63154
LOCUS R63154 481 bp mRNA EST 26-MAY-1995
DEFINITION yi01a11.r1 Soares placenta Nb2HP Homo sapiens cDNA clone IMAGE:137948 5', mRNA sequence.
ACCESSION R63154
VERSION R63154.1 GI:835033
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 481)
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and Wilson,R.
TITLE The WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Insert Size: 769
High qality sequence stops: 411
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 769 Std Error: 0.00
Seq primer: M13RP1
High quality sequence stop: 411.
FEATURES Location/Qualifiers
source 1. .481
/organism="Homo sapiens"
/db_xref="GDB:544291"
/db_xref="taxon:9606"
/clone="IMAGE:137948"
/clone_lib="Soares placenta Nb2HP"
/sex="Female"
/dev_stage="placenta obtained at birth (full term)"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: placenta; Vector: pT7T3D (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' AACTGGAAGAATTCGCGGCCGCAGGAATTTTTTTTTTTTTTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo. "
BASE COUNT 112 a 114 c 99 g 152 t 4 others
ORIGIN

Query Match 84.8%; Score 17.8; DB 188; Length 481;
Best Local Similarity 90.5%; Pred. No. 1.6e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 aagtgaactccctgctacctt 21
|||| | ||||||||||||||||
Db 11 AAGTTAGCTCCCTGCTACCTT 31

RESULT 6
AQ882343
LOCUS AQ882343 498 bp DNA GSS 09-NOV-1999
DEFINITION HS_5382_A1_B03_T7C RPCI-11 Human Male BAC Library Homo sapiens genomic clone Plate=9150 Col=5 Row=C, DNA sequence.
ACCESSION AQ882343
VERSION AQ882343.1 GI:6313810
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 498)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm) or from Resear h Genetics (info@resgen.com). BAC end Web Server: http://www.htsc.washington.edu
Plate: 9150 row: C column: 5
Seq primer: T7
Class: BAC ends
High quality sequence stop: 498.
FEATURES Location/Qualifiers
source 1. .498
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=9150 Col=5 Row=C"
/clone_lib="RPCI-11 Human Male BAC Library"
/sex="male"
/note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI; Male blood DNA was isolated from one randomly chosen donor and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACe3.6 vector at EcoRI sites"
BASE COUNT 155 a 111 c 118 g 111 t 3 others
ORIGIN

Query Match 84.8%; Score 17.8; DB 235; Length 498;
Best Local Similarity 90.5%; Pred. No. 1.6e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy      1 aagtgaactccctgctaccttt 21
          ||||||||||||| || ||||
Db     74 AAGTGAACTCCCTCCTGCCTT 94


RESULT   7
AQ185931/c
LOCUS       AQ185931      510 bp    DNA             GSS       01-NOV-1998
DEFINITION  HS_3065_A1_D09_MF CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=3065 Col=17 Row=G, DNA sequence.
ACCESSION   AQ185931
VERSION     AQ185931.1  GI:3585373
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 510)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Sequence Tagged Connector
            Plate: 3065  row: G  column: 17
            Class: BAC ends
            High quality sequence stop: 510.
FEATURES             Location/Qualifiers
     source          1. .510
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=3065 Col=17 Row=G"
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
                     /sex="male"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E-Coli DH10B"
BASE COUNT      157 a     98 c    100 g    153 t      2 others
ORIGIN


  Query Match             84.8%;  Score 17.8;  DB 225;  Length 510;
  Best Local Similarity   90.5%;  Pred. No. 1.6e+02;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 aagtgaactccctgctaccttt 21
          || ||||| |||||||||||||
Db    159 AAATGAACACCCTGCTACCTT 139


RESULT   8
AQ551067
LOCUS       AQ551067      608 bp    DNA             GSS       28-MAY-1999
DEFINITION  RPCI-11-416M22.TJ RPCI-11 Homo sapiens genomic clone RPCI-11-416M22
            , DNA sequence.
ACCESSION   AQ551067
VERSION     AQ551067.1  GI:4910244
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 608)
  AUTHORS   Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter
            ,J.C.
  TITLE     Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
            Map Building
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: hbe@tigr.org
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
            Research Genet cs (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .608
                     /organism="Homo sapiens"
                     /db_xref="GDB:7659669"
                     /db_xref="taxon:9606"
                     /clone="RPCI-11-416M22"
                     /clone_lib="RPCI-11"
                     /sex="Male"
                     /cell_type="Lymphocytes"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     RPCI11 Human Male BAC Library"
BASE COUNT      179 a    152 c    134 g    143 t
ORIGIN


  Query Match             84.8%;  Score 17.8;  DB 230;  Length 608;
  Best Local Similarity   90.5%;  Pred. No. 1.7e+02;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 aagtgaactccctgctaccttt 21
          |||| | |||||||||||||||
Db    376 AAGTTAGCTCCCTGCTACCTT 396


RESULT   9
AV553454/c
LOCUS       AV553454      471 bp    mRNA            EST       07-SEP-2000
DEFINITION  AV553454 Arabidopsis thaliana roots Columbia Arabidopsis thaliana
            cDNA clone RZ63b05R 5', mRNA sequence.
ACCESSION   AV553454
VERSION     AV553454.1  GI:8724867
KEYWORDS    EST.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 471)
  AUTHORS   Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
  TITLE     A large scale analysis of cDNA in Arabidopsis thaliana: Generation
            of 12,028 non-redundant expressed sequence tags from normalized and
            size-selected cDNA libraries
  JOURNAL   DNA Res. 7, 175-180 (2000)
  MEDLINE   20363093
COMMENT     Contact: Erika Asamizu
            The First Laboratory for Plant Gene Research
            Kazusa DNA Research Institute
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES             Location/Qualifiers
     source          1. .471
                     /organism="Arabidopsis thaliana"
                     /strain="Columbia"
                     /db_xref="taxon:3702"
```

/clone="RZ63b05R"
/clone_lib="Arabidopsis thaliana roots Columbia"
/tissue_type="roots"
/note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2: XhoI"
BASE COUNT 129 a 100 c 110 g 132 t
ORIGIN

Query Match 81.0%; Score 17; DB 31; Length 471;
Best Local Similarity 100.0%; Pred. No. 4.1e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 gaactccctgctacctt 21
||||||||||||||||||
Db 350 GAACTCCCTGCTACCTT 334

RESULT 10
AW381670/c
LOCUS AW381670 193 bp mRNA EST 04-FEB-2000
DEFINITION QV0-HT0310-061299-069-g09 HT0310 Homo sapiens cDNA, mRNA sequence.
ACCESSION AW381670
VERSION AW381670.1 GI:6886329
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 193)
AUTHORS HCGP http://www.ludwig.org.br/ORESTES.
TITLE The FAPESP/LICR Human Cancer Genome Project
JOURNAL Unpublished (1999)
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=QV0&t2=QV0-HT0310-061299-069-g09&t3=1999-12-06&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 17
High quality sequence stop: 193.
FEATURES Location/Qualifiers
source 1. .193
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="HT0310"
/dev_stage="Adult"
/note="Organ: head_neck; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
BASE COUNT 49 a 46 c 47 g 51 t
ORIGIN

Query Match 80.0%; Score 16.8; DB 115; Length 193;
Best Local Similarity 90.0%; Pred. No. 4.4e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 agtgaactccctgctacctt 21
||||||||||| | ||||||
Db 157 AGTGAACTCCCAGGTACCTT 138

RESULT 11
AA500281
LOCUS AA500281 304 bp mRNA EST 01-JUL-1997
DEFINITION vi97f08.r1 Barstead mouse pooled organs MPLRB4 Mus musculus cDNA clone IMAGE:920199 5', mRNA sequence.
ACCESSION AA500281
VERSION AA500281.1 GI:2235248
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 304)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
TITLE The WashU-HHMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:532415
Seq primer: -28m13 rev2 ET from Amersham
High quality sequence stop: 295.
FEATURES Location/Qualifiers
source 1. .304
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:920199"
/clone_lib="Barstead mouse pooled organs MPLRB4"
/sex="mixed"
/tissue_type="pooled organs"
/dev_stage="7 day"
/lab_host="DH10B"
/note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: EcoRI; Site_2: NotI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACGAATCTGAAGTGGGAGCGGCCGCCCTTTTTTTTTTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors [GTTGGATTCGGTACC], digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library constructed by Bob Barstead."
BASE COUNT 74 a 84 c 76 g 70 t
ORIGIN

Query Match 80.0%; Score 16.8; DB 8; Length 304;
Best Local Similarity 90.0%; Pred. No. 4.8e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 aagtgaactccctgctacct 20
|||||||||||||| | |||
Db 152 AAGTGAACTCCCTGGTTCCT 171

RESULT 12
AA560193
LOCUS AA560193 326 bp mRNA EST 18-AUG-1997
DEFINITION vl20g01.r1 Stratagene mouse Tcell 937311 Mus musculus cDNA clone IMAGE:972816 5', mRNA sequence.
ACCESSION AA560193
VERSION AA560193.1 GI:2331658
KEYWORDS EST.

```
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 326)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:553544
            Seq primer: -28m13 rev1 ET from Amersham
            High quality sequence stop: 294.
FEATURES             Location/Qualifiers
     source          1. .326
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:972816"
                     /clone_lib="Stratagene mouse Tcell 937311"
                     /tissue_type="Tcell"
                     /dev_stage="M30 CD4+ cells"
                     /lab_host="SOLR (kanamycin resistant)"
                     /note="Organ: blood; Vector: pBluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI; Cloned unidirectionally.  Primer:
                     Oligo dT. M30 CD4+ cells. Average insert size: 1.0 kb;
                     Uni-ZAP XR Vector; ~5' adaptor sequence: 5' GAATTCGGCACGAG
                     3' ~3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTTTT 3'"
BASE COUNT       71 a     89 c     89 g     76 t      1 others
ORIGIN
```

```
  Query Match           80.0%;  Score 16.8;  DB 9;  Length 326;
  Best Local Similarity 90.0%;  Pred. No. 4.8e+02;
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 aagtgaactccctgctacct 20
           |||||||||||||| | |||
Db     103 AAGTGAACTCCCTGGTGCCT 122
```

```
RESULT  13
AA620205
LOCUS       AA620205      410 bp    mRNA            EST       09-OCT-1997
DEFINITION  vo64a06.r1 Soares_mammary_gland_NbMMG Mus musculus cDNA clone
            IMAGE:1054642 5', mRNA sequence.
ACCESSION   AA620205
VERSION     AA620205.1  GI:2524081
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 410)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:586218
            Seq primer: -28m13 rev2 ET from Amersham
            High quality sequence stop: 400.
FEATURES             Location/Qualifiers
     source          1. .410
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:1054642"
                     /clone_lib="Soares_mammary_gland_NbMMG"
                     /sex="male"
                     /tissue_type="mammary gland"
                     /dev_stage="4 weeks"
                     /lab_host="DH10B"
                     /note="Organ: mammary gland; Vector: pT7T3D-Pac (Pharmacia
                     ) with a modified polylinker; Site_1: Not I; Site_2: Eco
                     RI; 1st strand cDNA was primed with a Not I - oligo(dT)
                     primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGAATGGTTTTTTTTTTTTTTTTTTTT
                     T 3']; double-stranded cDNA was ligated to Eco RI
                     adaptors (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of the modified pT7T3 vector.
                     RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
                     constructed and normalized by Bento Soares and M.Fatima
                     Bonaldo."
BASE COUNT       95 a     116 c     115 g      84 t
ORIGIN
```

```
  Query Match           80.0%;  Score 16.8;  DB 9;  Length 410;
  Best Local Similarity 90.0%;  Pred. No. 5e+02;
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 aagtgaactccctgctacct 20
           |||||||| |||||  |||||
Db       8 AAGTGAACGCCCTGATACCT 27
```

```
RESULT  14
AI615361
LOCUS       AI615361      416 bp    mRNA            EST       21-APR-1999
DEFINITION  vi97f08.y1 Barstead mouse pooled organs MPLRB4 Mus musculus cDNA
            clone IMAGE:920199 5' similar to TR:O75935 O75935 DYNACTIN SUBUNIT.
            ;, mRNA sequence.
ACCESSION   AI615361
VERSION     AI615361.1  GI:4624528
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 416)
  AUTHORS   Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
            Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person
            ,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter
            ,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
            Waterston,R. and Wilson,R.
  TITLE     The WashU-NCI Mouse EST Project 1999
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Marra M/WashU-NCI Mouse EST Project 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
```

IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:532415
This read is a RESEQUENCE of a previously sequenced mouse clone
This read has been verified (found to hit its original self in the correct orientation)
Putative full length read
vector to vector length is 646
Seq primer: -40RP from Gibco
High quality sequence stop: 411.

FEATURES Location/Qualifiers
source 1. .416
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:920199"
/clone_lib="Barstead mouse pooled organs MPLRB4"
/sex="mixed"
/tissue_type="pooled organs"
/dev_stage="7 day"
/lab_host="DH10B"
/note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: EcoRI; Site_2: NotI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACGAATCTGAAGTGGGAGCGGCCGCCCTTTTTTTTTTTTTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors [GTTGGATTCGGTACC], digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library constructed by Bob Barstead."

BASE COUNT 105 a 111 c 107 g 93 t
ORIGIN

Query Match 80.0%; Score 16.8; DB 22; Length 416;
Best Local Similarity 90.0%; Pred. No. 5e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 aagtgaactccctgctacct 20
||||||||||||||| | |||
Db 154 AAGTGAACTCCCTGGTGCCT 173

RESULT 15
AA616429
LOCUS AA616429 439 bp mRNA EST 07-OCT-1997
DEFINITION vo09e02.r1 Barstead mouse myotubes MPLRB5 Mus musculus cDNA clone IMAGE:1049402 5', mRNA sequence.
ACCESSION AA616429
VERSION AA616429.1 GI:2503634
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 439)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
TITLE The WashU-HHMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box B501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:580978
Seq primer: -28m13 rev2 ET from Amersham.

FEATURES Location/Qualifiers
source 1. .439
/organism="Mus musculus"
/strain="C3H"
/db_xref="taxon:10090"
/clone="IMAGE:1049402"
/clone_lib="Barstead mouse myotubes MPLRB5"
/cell_line="C2C12"
/lab_host="DH10B"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: EcoRI; Site_2: NotI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACGAATCTGAAGTGGGAGCGGCCGCCCTTTTTTTTTTTTTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors [AATTCGGATCCTTG], digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library constructed by Bob Barstead. The C2C12 cell line (available from ATCC, catalog # CRL-1772) differentiates rapidly, forming contractile myotubes and producing characteristic muscle proteins."

BASE COUNT 105 a 108 c 121 g 105 t
ORIGIN

Query Match 80.0%; Score 16.8; DB 9; Length 439;
Best Local Similarity 90.0%; Pred. No. 5.1e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 aagtgaactccctgctacct 20
||||||||||||||| | |||
Db 214 AAGTGAACTCCCTGGTGCCT 233

Search completed: July 25, 2001, 04:54:49
Job time: 10422 sec

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: July 25, 2001, 05:18:51 ; Search time 117.39 Seconds
(without alignments)
33.141 Million cell updates/sec

Title: US-09-142-095-2
Perfect score: 21
Sequence: 1 ccactgggatcaacagtatct 21

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 317530 seqs, 92630169 residues

Total number of hits satisfying chosen parameters: 635060

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	21	100.0	1190	5	PCT-US92-00282-18	Sequence 18, Appl
c	2	21	100.0	2336	5	PCT-US92-00282-1	Sequence 1, Appli
	3	15.8	75.2	1073	4	US-08-960-780-37	Sequence 37, Appl
	4	15.8	75.2	1073	4	US-09-073-898-37	Sequence 37, Appl
	5	14.8	70.5	831	2	US-08-403-853-17	Sequence 17, Appl
c	6	14.8	70.5	832	1	US-08-473-496-1	Sequence 1, Appli
c	7	14.8	70.5	832	2	US-08-454-028-3	Sequence 3, Appli
c	8	14.8	70.5	832	5	PCT-US94-05388-3	Sequence 3, Appli
c	9	14.8	70.5	832	5	PCT-US95-09121-1	Sequence 1, Appli
c	10	14.8	70.5	832	5	PCT-US96-07496-3	Sequence 3, Appli
c	11	14.8	70.5	3225	1	US-08-698-551-13	Sequence 13, Appl
c	12	14.8	70.5	3225	2	US-08-602-228-13	Sequence 13, Appl
c	13	14.8	70.5	3225	2	US-08-533-901B-13	Sequence 13, Appl
c	14	14.8	70.5	3225	2	US-08-839-032A-13	Sequence 13, Appl
c	15	14.8	70.5	3225	2	US-08-839-031A-13	Sequence 13, Appl
c	16	14.8	70.5	3225	5	PCT-US95-12724-13	Sequence 13, Appl
	17	14.6	69.5	87350	3	US-08-781-891-79	Sequence 79, Appl
c	18	14.2	67.6	20	4	US-08-669-378-16	Sequence 16, Appl
c	19	14.2	67.6	330	1	US-08-558-735-19	Sequence 19, Appl
c	20	14.2	67.6	330	4	US-08-906-480-19	Sequence 19, Appl
c	21	14.2	67.6	333	1	US-08-558-735-18	Sequence 18, Appl
c	22	14.2	67.6	333	4	US-08-906-480-18	Sequence 18, Appl
	23	14.2	67.6	336	2	US-08-888-366-17	Sequence 17, Appl
c	24	14.2	67.6	343	1	US-08-320-559-36	Sequence 36, Appl
c	25	14.2	67.6	343	3	US-08-545-860D-36	Sequence 36, Appl
c	26	14.2	67.6	343	5	PCT-US94-04496-36	Sequence 36, Appl
c	27	14.2	67.6	345	1	US-08-558-735-3	Sequence 3, Appli
c	28	14.2	67.6	345	1	US-08-558-735-17	Sequence 17, Appl
c	29	14.2	67.6	345	4	US-08-906-480-3	Sequence 3, Appli
c	30	14.2	67.6	345	4	US-08-906-480-17	Sequence 17, Appl
c	31	14.2	67.6	355	1	US-08-558-735-16	Sequence 16, Appl
c	32	14.2	67.6	355	4	US-08-906-480-16	Sequence 16, Appl
	33	14.2	67.6	428	3	US-08-589-939-4	Sequence 4, Appli
c	34	14.2	67.6	532	1	US-08-558-735-1	Sequence 1, Appli
c	35	14.2	67.6	532	4	US-08-906-480-1	Sequence 1, Appli
	36	14.2	67.6	780	3	US-09-027-449-54	Sequence 54, Appl
	37	14.2	67.6	780	3	US-09-027-449-58	Sequence 58, Appl
	38	14.2	67.6	780	3	US-09-027-449-65	Sequence 65, Appl
	39	14.2	67.6	780	4	US-08-804-444A-54	Sequence 54, Appl
	40	14.2	67.6	780	4	US-08-804-444A-58	Sequence 58, Appl
	41	14.2	67.6	780	4	US-09-026-985-54	Sequence 54, Appl
	42	14.2	67.6	780	4	US-09-026-985-58	Sequence 58, Appl
	43	14.2	67.6	780	4	US-09-026-985-65	Sequence 65, Appl
	44	14.2	67.6	1422	4	US-08-387-117-8	Sequence 8, Appli
	45	14.2	67.6	1905	4	US-08-387-117-7	Sequence 7, Appli

ALIGNMENTS

```
RESULT 1
PCT-US92-00282-18/c
; Sequence 18, Application PC/TUS9200282
;  GENERAL INFORMATION:
;    APPLICANT:  OWENS, IDA S.
;    APPLICANT:  RITTER, JOSEPH K.
;    TITLE OF INVENTION:  THE GENETIC LOCUS UGT1 AND A MUTATION
;    TITLE OF INVENTION:  THEREIN.
;    NUMBER OF SEQUENCES:  40
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN DARBY & CUSHMAN
;      STREET:  1615 L STREET, N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20036-5601
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US92/00282
;      FILING DATE:  19920110
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SCOTT, WATSON T.
;      REGISTRATION NUMBER:  26581
;      REFERENCE/DOCKET NUMBER:  91532-PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-861-3000
;      TELEFAX:  202-822-0944
;      TELEX:  6714627 CUSH
;  INFORMATION FOR SEQ ID NO:  18:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1190 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
PCT-US92-00282-18


  Query Match          100.0%;  Score 21;  DB 5;  Length 1190;
  Best Local Similarity  100.0%;  Pred. No. 0.075;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ccactgggatcaacagtatct 21
          |||||||||||||||||||||
Db    190 CCACTGGGATCAACAGTATCT 170
```

```
RESULT   2
PCT-US92-00282-1/c
; Sequence 1, Application PC/TUS9200282
;  GENERAL INFORMATION:
;    APPLICANT:  OWENS, IDA S.
;    APPLICANT:  RITTER, JOSEPH K.
;    TITLE OF INVENTION:  THE GENETIC LOCUS UGT1 AND A MUTATION
;    TITLE OF INVENTION:  THEREIN.
;    NUMBER OF SEQUENCES:  40
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN DARBY & CUSHMAN
;      STREET:  1615 L STREET, N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20036-5601
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US92/00282
;      FILING DATE:  19920110
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SCOTT, WATSON T.
;      REGISTRATION NUMBER:  26581
;      REFERENCE/DOCKET NUMBER:  91532-PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-861-3000
;      TELEFAX:  202-822-0944
;      TELEX:  6714627 CUSH
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2336 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
PCT-US92-00282-1


  Query Match           100.0%;  Score 21;  DB 5;  Length 2336;
  Best Local Similarity  100.0%;  Pred. No. 0.082;
  Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ccactgggatcaacagtatct 21
           |||||||||||||||||||||
Db     106 CCACTGGGATCAACAGTATCT 86


RESULT   3
US-08-960-780-37
; Sequence 37, Application US/08960780
; Patent No. 6204435
;  GENERAL INFORMATION:
;    APPLICANT:  Feitelson, Jerald S.
;    APPLICANT:  Schnepf, H. Ernest
;    APPLICANT:  Narva, Kenneth E.
;    APPLICANT:  Stockhoff, Brian A.
;    APPLICANT:  Schmeits, James
;    APPLICANT:  Loewer, David
;    APPLICANT:  Dullum, Charles Joseph
;    APPLICANT:  Muller-Cohn, Judy
;    APPLICANT:  Stamp, Lisa
;    TITLE OF INVENTION:  No. 6204435el Pesticidal Toxins and Nucleotide
;    TITLE OF INVENTION:  Sequences Which Encode These Toxins
;    NUMBER OF SEQUENCES:  134
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Saliwanchik, Lloyd & Saliwanchik
;      STREET:  2421 N.W. 41st Street, Suite A-1
;      CITY:  Gainesville
;      STATE:  FL
;      COUNTRY:  US
;      ZIP:  32606-6669
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/960,780
;      FILING DATE:  30-OCT-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/029,848
;      FILING DATE:  30-OCT-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Saliwanchik, David R.
;      REGISTRATION NUMBER:  31,794
;      REFERENCE/DOCKET NUMBER:  MA-708
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  352-375-8100
;      TELEFAX:  352-372-5800
;  INFORMATION FOR SEQ ID NO:  37:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1073 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      INDIVIDUAL ISOLATE:  196F3
US-08-960-780-37


  Query Match           75.2%;  Score 15.8;  DB 4;  Length 1073;
  Best Local Similarity  89.5%;  Pred. No. 29;
  Matches   17;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       3 actgggatcaacagtatct 21
           |||||||  ||||||||||
Db     838 ACTGGGAAAAACAGTATCT 856


RESULT   4
US-09-073-898-37
; Sequence 37, Application US/09073898
; Patent No. 6242669
;  GENERAL INFORMATION:
;    APPLICANT:  Feitelson, Jerald S.
;    APPLICANT:  Schnepf, H. Ernest
;    APPLICANT:  Narva, Kenneth E.
;    APPLICANT:  Stockhoff, Brian A.
;    APPLICANT:  Schmeits, James
;    APPLICANT:  Loewer, David
;    APPLICANT:  Dullum, Charles Joseph
;    APPLICANT:  Muller-Cohn, Judy
;    APPLICANT:  Stamp, Lisa
;    APPLICANT:  Morrill, George
;    APPLICANT:  Finstad-Lee, Stacey
;    TITLE OF INVENTION:  No. 6242669el Pesticidal Toxins and Nucleotide
;    TITLE OF INVENTION:  Sequences Which Encode These Toxins
;    NUMBER OF SEQUENCES:  144
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Saliwanchik, Lloyd & Saliwanchik
;      STREET:  2421 N.W. 41st Street, Suite A-1
;      CITY:  Gainesville
;      STATE:  FL
;      COUNTRY:  US
;      ZIP:  32606-6669
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
```

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/073,898
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/029,848
; FILING DATE: 30-OCT-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/960,780
; FILING DATE: 30-OCT-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Sanders, Jay M.
; REGISTRATION NUMBER: 39,355
; REFERENCE/DOCKET NUMBER: MA-708C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 352-375-8100
; TELEFAX: 352-372-5800
; INFORMATION FOR SEQ ID NO: 37:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1073 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: 196F3
US-09-073-898-37

Query Match 75.2%; Score 15.8; DB 4; Length 1073;
Best Local Similarity 89.5%; Pred. No. 29;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 actgggatcaacagtatct 21
||||||| |||||||||
Db 838 ACTGGGAAAAACAGTATCT 856

RESULT 5
US-08-403-853-17
; Sequence 17, Application US/08403853
; Patent No. 5844094
; GENERAL INFORMATION:
; APPLICANT: HUDSON, Peter J.
; APPLICANT: LAH, Maria
; APPLICANT: KORRT, Alex A.
; APPLICANT: IRVING, Robert A.
; APPLICANT: ATWELL, John L.
; APPLICANT: MALBY, Robyn L.
; APPLICANT: POWER, Barbara E.
; APPLICANT: COLMAN, Peter M.
; TITLE OF INVENTION: TARGET BINDING POLYPEPTIDE
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/403,853
; FILING DATE: 30-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/AU93/00491
; FILING DATE: 24-SEP-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: AU PL 4973
; FILING DATE: 25-SEP-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 16786/189/CHAC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 831 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 1..819
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..819
US-08-403-853-17

Query Match 70.5%; Score 14.8; DB 2; Length 831;
Best Local Similarity 88.9%; Pred. No. 88;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 actgggatcaacagtatc 20
||||| |||||||| |||
Db 575 ACTGGTATCAACAGAATC 592

RESULT 6
US-08-473-496-1/c
; Sequence 1, Application US/08473496
; Patent No. 5700660
; GENERAL INFORMATION:
; APPLICANT: Jack L. Leonard
; APPLICANT: Peter E. Newburger
; TITLE OF INVENTION: POSITIONAL CONTROL OF SELENIUM INSERTION
; TITLE OF INVENTION: IN POLYPEPTIDES FOR X-RAY CRYSTALLOGRAPHY
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 50Z or 55SX
; OPERATING SYSTEM: MS-DOS (Version 5.0)
; SOFTWARE: WordPerfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/473,496
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/277,492
; FILING DATE: 19-JUL-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/066,680
; FILING DATE: 24-MAY-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Fasse, J. Peter
; REGISTRATION NUMBER: 32,983
; REFERENCE/DOCKET NUMBER: 04020/078001

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 832 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
US-08-473-496-1

Query Match 70.5%; Score 14.8; DB 1; Length 832;
Best Local Similarity 88.9%; Pred. No. 88;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 actgggatcaacagtatc 20
||||||||||||||| | |
Db 768 ACTGGGATCAACAGGACC 751

RESULT 7
US-08-454-028-3/c
; Sequence 3, Application US/08454028
; Patent No. 5849520
; GENERAL INFORMATION:
; APPLICANT: Jack L. Leonard
; APPLICANT: Peter E. Newburger
; TITLE OF INVENTION: POST-TRANSCRIPTIONAL GENE REGULATION BY
; TITLE OF INVENTION: SELENIUM
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 50Z or 55SX
; OPERATING SYSTEM: MS-DOS (Version 5.0)
; SOFTWARE: WordPerfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/454,028
; FILING DATE: May 30, 1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/066,680
; FILING DATE: May 24, 1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Fasse, J. Peter
; REGISTRATION NUMBER: 32,983
; REFERENCE/DOCKET NUMBER: 04020/075001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 832 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
US-08-454-028-3

Query Match 70.5%; Score 14.8; DB 2; Length 832;
Best Local Similarity 88.9%; Pred. No. 88;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 actgggatcaacagtatc 20
||||||||||||||| | |
Db 768 ACTGGGATCAACAGGACC 751

RESULT 8
PCT-US94-05388-3/c
; Sequence 3, Application PC/TUS9405388
; GENERAL INFORMATION:
; APPLICANT: Jack L. Leonard
; APPLICANT: Peter E. Newburger
; TITLE OF INVENTION: POST-TRANSCRIPTIONAL GENE
; TITLE OF INVENTION: REGULATION BY TRACE
; NUMBER OF SEQUENCES: 28
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 50Z or 55SX
; OPERATING SYSTEM: MS-DOS (Version 5.0)
; SOFTWARE: WordPerfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/05388
; FILING DATE: 16 May 1994
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/066,680
; FILING DATE: May 24, 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Paul T.
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 04020/026001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 832
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
PCT-US94-05388-3

Query Match 70.5%; Score 14.8; DB 5; Length 832;
Best Local Similarity 88.9%; Pred. No. 88;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 actgggatcaacagtatc 20
||||||||||||||| | |
Db 768 ACTGGGATCAACAGGACC 751

RESULT 9
PCT-US95-09121-1/c
; Sequence 1, Application PC/TUS9509121
; GENERAL INFORMATION:
; APPLICANT: University of Massachusetts
; APPLICANT: Medical School
; TITLE OF INVENTION: POSITIONAL CONTROL OF SELENIUM
; TITLE OF INVENTION: INSERTION IN POLYPEPTIDES FOR
; TITLE OF INVENTION: X-RAY CRYSTALLOGRAPHY
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.

```
;       STREET:  225 Franklin Street
;       CITY:  Boston
;       STATE:  Massachusetts
;       COUNTRY:  U.S.A.
;       ZIP:  02110-2804
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
;       COMPUTER:  IBM PS/2 Model 50Z or
;       COMPUTER:  55SX
;       OPERATING SYSTEM:  MS-DOS (Version 5.0)
;       SOFTWARE:  WordPerfect (Version
;       SOFTWARE:  5.1)
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/US95/09121
;       FILING DATE:  19 July 1995
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/277,492
;       FILING DATE:  19 July 1994
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/473,496
;       FILING DATE:  7 June 1995
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Clark, Paul T.
;       REGISTRATION NUMBER:  30,162
;       REFERENCE/DOCKET NUMBER:  04020/078WO1
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (617) 542-5070
;       TELEFAX:  (617) 542-8906
;       TELEX:  200154
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  832 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA
PCT-US95-09121-1


  Query Match             70.5%;  Score 14.8;  DB 5;  Length 832;
  Best Local Similarity   88.9%;  Pred. No. 88;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      3 actgggatcaacagtatc 20
          ||||||||||||||| | |
Db    768 ACTGGGATCAACAGGACC 751


RESULT  10
PCT-US96-07496-3/c
; Sequence 3, Application PC/TUS9607496
;  GENERAL INFORMATION:
;     APPLICANT:  University of Massachusetts Medical Center
;     TITLE OF INVENTION:  POST-TRANSCRIPTIONAL GENE REGULATION BY
;     TITLE OF INVENTION:  SELENIUM
;     NUMBER OF SEQUENCES:  33
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Fish & Richardson P.C.
;       STREET:  225 Franklin Street
;       CITY:  Boston
;       STATE:  Massachusetts
;       COUNTRY:  U.S.A.
;       ZIP:  02110-2804
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
;       COMPUTER:  IBM PS/2 Model 50Z or 55SX
;       OPERATING SYSTEM:  MS-DOS (Version 5.0)
;       SOFTWARE:  WordPerfect (Version 5.1)
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/US96/07496
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/454,028
;       FILING DATE:  May 30, 1995
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/066,680
;       FILING DATE:  May 24, 1993
;       CLASSIFICATION:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Fasse, J. Peter
;       REGISTRATION NUMBER:  32,983
;       REFERENCE/DOCKET NUMBER:  04020/075WO1
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (617) 542-5070
;       TELEFAX:  (617) 542-8906
;       TELEX:  200154
;   INFORMATION FOR SEQ ID NO:  3:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  832 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA
PCT-US96-07496-3


  Query Match             70.5%;  Score 14.8;  DB 5;  Length 832;
  Best Local Similarity   88.9%;  Pred. No. 88;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      3 actgggatcaacagtatc 20
          ||||||||||||||| | |
Db    768 ACTGGGATCAACAGGACC 751


RESULT  11
US-08-698-551-13/c
; Sequence 13, Application US/08698551
; Patent No. 5712381
;  GENERAL INFORMATION:
;     APPLICANT:  Lin, Lih-Ling
;     APPLICANT:  Chen, Jennifer H.
;     APPLICANT:  Schievella, Andrea
;     APPLICANT:  Graham, James
;     TITLE OF INVENTION:  NOVEL TNF RECEPTOR DEATH DOMAIN LIGAND
;     TITLE OF INVENTION:  PROTEINS AND INHIBITORS OF LIGAND BINDING
;     NUMBER OF SEQUENCES:  18
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Genetics Institute, Inc.
;       STREET:  87 CambridgePark Drive
;       CITY:  Cambridge
;       STATE:  Massachusetts
;       COUNTRY:  USA
;       ZIP:  02140
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/698,551
;       FILING DATE:
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Brown, Scott A,
;       REGISTRATION NUMBER:  32,724
;       REFERENCE/DOCKET NUMBER:  GI5232D
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (617) 498-8224
;       TELEFAX:  (617) 876-5851
;   INFORMATION FOR SEQ ID NO:  13:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  3225 base pairs
```

```
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     HYPOTHETICAL:  NO
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  3..2846
US-08-698-551-13


  Query Match             70.5%;  Score 14.8;  DB 1;  Length 3225;
  Best Local Similarity   88.9%;  Pred. No. 1e+02;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       4 ctgggatcaacagtatct 21
           |||||  ||||||||||
Db     504 CTGGGTCCAACAGTATCT 487


RESULT  12
US-08-602-228-13/c
; Sequence 13, Application US/08602228
; Patent No. 5843675
;  GENERAL INFORMATION:
;    APPLICANT:  Lin, Lih-Ling
;    APPLICANT:  Chen, Jennifer H.
;    APPLICANT:  Schievella, Andrea
;    APPLICANT:  Graham, James
;    TITLE OF INVENTION:  NOVEL TNF RECEPTOR DEATH DOMAIN LIGAND
;    TITLE OF INVENTION:  PROTEINS AND INHIBITORS OF LIGAND BINDING
;    NUMBER OF SEQUENCES:  16
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genetics Institute, Inc.
;      STREET:  87 CambridgePark Drive
;      CITY:  Cambridge
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02140
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/602,228
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Brown, Scott A,
;      REGISTRATION NUMBER:  32,724
;      REFERENCE/DOCKET NUMBER:  GI5232C
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 498-8224
;      TELEFAX:  (617) 876-5851
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3225 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  3..2846
US-08-602-228-13


  Query Match             70.5%;  Score 14.8;  DB 2;  Length 3225;
  Best Local Similarity   88.9%;  Pred. No. 1e+02;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       4 ctgggatcaacagtatct 21
           |||||  ||||||||||
Db     504 CTGGGTCCAACAGTATCT 487


RESULT  13
US-08-533-901B-13/c
; Sequence 13, Application US/08533901B
; Patent No. 5852173
;  GENERAL INFORMATION:
;    APPLICANT:  Lin, Lih-Ling
;    APPLICANT:  Chen, Jennifer H.
;    APPLICANT:  Schievella, Andrea
;    APPLICANT:  Graham, James
;    TITLE OF INVENTION:  NOVEL TNF RECEPTOR DEATH DOMAIN LIGAND
;    TITLE OF INVENTION:  PROTEINS AND INHIBITORS OF LIGAND BINDING
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genetics Institute, Inc.
;      STREET:  87 CambridgePark Drive
;      CITY:  Cambridge
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02140
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/533,901B
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Brown, Scott A,
;      REGISTRATION NUMBER:  32,724
;      REFERENCE/DOCKET NUMBER:  GI5232
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 498-8224
;      TELEFAX:  (617) 876-5851
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3225 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  3..2846
US-08-533-901B-13


  Query Match             70.5%;  Score 14.8;  DB 2;  Length 3225;
  Best Local Similarity   88.9%;  Pred. No. 1e+02;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       4 ctgggatcaacagtatct 21
           |||||  ||||||||||
Db     504 CTGGGTCCAACAGTATCT 487


RESULT  14
US-08-839-032A-13/c
; Sequence 13, Application US/08839032A
; Patent No. 5891675
;  GENERAL INFORMATION:
;    APPLICANT:  Lin, Lih-Ling
;    APPLICANT:  Chen, Jennifer H.
;    APPLICANT:  Schievella, Andrea
;    APPLICANT:  Graham, James
;    TITLE OF INVENTION:  NOVEL TNF RECEPTOR DEATH DOMAIN LIGAND PROTEINS
```

```
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genetics Institute, Inc.
;      STREET:  87 CambridgePark Drive
;      CITY:  Cambridge
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02140
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/839,032A
;      FILING DATE:
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Brown, Scott A,
;      REGISTRATION NUMBER:  32,724
;      REFERENCE/DOCKET NUMBER:  GI5232DDIV
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 498-8224
;      TELEFAX:  (617) 876-5851
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3225 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  3..2846
US-08-839-032A-13


  Query Match           70.5%;  Score 14.8;  DB 2;  Length 3225;
  Best Local Similarity  88.9%;  Pred. No. 1e+02;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       4 ctgggatcaacagtatct 21
           |||||  ||||||||||
Db     504 CTGGGTCCAACAGTATCT 487


RESULT  15
US-08-839-031A-13/c
; Sequence 13, Application US/08839031A
; Patent No. 5948638
;  GENERAL INFORMATION:
;    APPLICANT:  Lin, Lih-Ling
;    APPLICANT:  Chen, Jennifer H.
;    APPLICANT:  Schievella, Andrea
;    APPLICANT:  Graham, James
;    TITLE OF INVENTION:  NOVEL TNF RECEPTOR DEATH DOMAIN LIGAND
;    TITLE OF INVENTION:  PROTEINS
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genetics Institute, Inc.
;      STREET:  87 CambridgePark Drive
;      CITY:  Cambridge
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02140
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/839,031A
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sprunger, Suzanne A.,
;      REGISTRATION NUMBER:  41,323
;      REFERENCE/DOCKET NUMBER:  GI5232BDIV
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 498-8284
;      TELEFAX:  (617) 876-5851
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3225 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  3..2846
US-08-839-031A-13


  Query Match           70.5%;  Score 14.8;  DB 2;  Length 3225;
  Best Local Similarity  88.9%;  Pred. No. 1e+02;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       4 ctgggatcaacagtatct 21
           |||||  ||||||||||
Db     504 CTGGGTCCAACAGTATCT 487


Search completed:  July 25, 2001, 05:18:52
Job time:  9215 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: July 25, 2001, 04:54:49 ; Search time 2762.24 Seconds
(without alignments)
71.866 Million cell updates/sec

Title: US-09-142-095-2
Perfect score: 21
Sequence: 1 ccactgggatcaacagtatct 21

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est33:*
26: gb_est34:*
27: gb_est35:*
28: gb_est36:*
29: gb_est37:*
30: gb_est38:*
31: gb_est39:*
32: gb_est40:*
33: em_estba:*
34: em_estfun:*
35: em_esthum1:*
36: em_esthum2:*
37: em_esthum3:*
38: em_esthum4:*
39: em_esthum5:*
40: em_esthum6:*
41: em_esthum7:*
42: em_esthum8:*
43: em_esthum9:*
44: em_esthum10:*
45: em_esthum11:*
46: em_esthum12:*
47: em_esthum13:*
48: em_esthum14:*
49: em_esthum15:*
50: em_esthum16:*
51: em_esthum17:*
52: em_esthum18:*
53: em_esthum19:*
54: em_esthum20:*
55: em_esthum21:*
56: em_esthum22:*
57: em_esthum23:*
58: em_esthum24:*
59: em_esthum25:*
60: em_esthum26:*
61: em_esthum27:*
62: em_esthum28:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estin5:*
68: em_estom1:*
69: em_estom2:*
70: em_estov1:*
71: em_estov2:*
72: em_estpl1:*
73: em_estpl2:*
74: em_estpl3:*
75: em_estpl4:*
76: em_estpl5:*
77: em_estpl6:*
78: em_estpl7:*
79: em_estpl8:*
80: em_estpl9:*
81: em_estpl10:*
82: em_estro1:*
83: em_estro2:*
84: em_estro3:*
85: em_estro4:*
86: em_estro5:*
87: em_estro6:*
88: em_estro7:*
89: em_estro8:*
90: em_estro9:*
91: em_estro10:*
92: em_estro11:*
93: em_estro12:*
94: em_estro13:*
95: em_estro14:*
96: em_estro15:*
97: em_estro16:*
98: em_estro17:*
99: em_estro18:*
100: em_estro19:*
101: em_estro20:*
102: gb_est25:*
103: gb_est26:*
104: gb_est27:*
105: gb_est28:*
106: gb_est29:*
107: gb_est30:*
108: gb_est31:*
109: gb_est32:*
110: gb_est41:*
111: gb_est42:*
112: gb_est43:*
113: gb_est44:*
114: gb_est45:*
115: gb_est46:*
116: gb_est47:*

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est75:*
137: gb_est76:*
138: gb_est77:*
139: gb_est78:*
140: gb_est79:*
141: gb_est80:*
142: gb_est81:*
143: gb_est82:*
144: gb_est83:*
145: gb_est84:*
146: gb_est85:*
147: gb_est86:*
148: gb_est87:*
149: gb_est88:*
150: gb_est89:*
151: gb_est90:*
152: gb_est99:*
153: gb_est100:*
154: gb_est101:*
155: gb_est102:*
156: gb_est103:*
157: gb_est104:*
158: gb_est105:*
159: gb_est106:*
160: gb_est67:*
161: gb_est68:*
162: gb_est69:*
163: gb_est70:*
164: gb_est71:*
165: gb_est72:*
166: gb_est73:*
167: gb_est74:*
168: gb_est91:*
169: gb_est92:*
170: gb_est93:*
171: gb_est94:*
172: gb_est95:*
173: gb_est96:*
174: gb_est97:*
175: gb_est98:*
176: em_esthum29:*
177: em_esthum30:*
178: em_esthum31:*
179: em_esthum32:*
180: em_esthum33:*
181: em_estom3:*
182: em_estpl11:*
183: em_estro21:*
184: em_estro22:*
185: em_estro23:*
186: em_htc:*
187: gb_est107:*
188: gb_est108:*
189: gb_est109:*
190: gb_est110:*
191: gb_est111:*
192: gb_htc:*
193: em_gss_fun:*
194: em_gss_hum1:*
195: em_gss_hum2:*
196: em_gss_hum3:*
197: em_gss_hum4:*
198: em_gss_hum5:*
199: em_gss_hum6:*
200: em_gss_hum7:*
201: em_gss_hum8:*
202: em_gss_hum9:*
203: em_gss_inv1:*
204: em_gss_inv2:*
205: em_gss_inv3:*
206: em_gss_other:*
207: em_gss_pln1:*
208: em_gss_pln2:*
209: em_gss_pro:*
210: em_gss_rod1:*
211: em_gss_rod2:*
212: em_gss_rod3:*
213: em_gss_rod4:*
214: em_gss_rod5:*
215: em_gss_vrt1:*
216: em_gss_vrt2:*
217: em_gss_vrt3:*
218: gb_gss1:*
219: gb_gss2:*
220: gb_gss3:*
221: gb_gss4:*
222: gb_gss5:*
223: gb_gss6:*
224: gb_gss7:*
225: gb_gss8:*
226: gb_gss9:*
227: gb_gss10:*
228: gb_gss11:*
229: gb_gss12:*
230: gb_gss13:*
231: gb_gss14:*
232: gb_gss15:*
233: gb_gss16:*
234: gb_gss17:*
235: gb_gss18:*
236: gb_gss19:*
237: gb_gss20:*
238: gb_gss21:*
239: gb_gss22:*
240: gb_gss23:*
241: gb_gss24:*
242: gb_gss25:*
243: gb_gss26:*
244: gb_gss27:*
245: gb_gss28:*
246: gb_gss29:*
247: gb_gss30:*
248: gb_gss31:*
249: gb_gss32:*
250: gb_gss33:*
251: gb_gss34:*
252: em_gss_inv4:*
253: em_gss_rod6:*
254: em_gss_rod7:*
255: em_gss_rod8:*
256: gb_gss35:*
257: gb_gss36:*
258: gb_gss37:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
c 1	21	100.0	327	189	T71061	T71061 yc50a04.r1
2	18.4	87.6	383	3	AA195421	AA195421 zr36h09.s
c 3	18.4	87.6	1078	221	CNS04CB0	AL284229 Tetraodon
c 4	17.8	84.8	297	169	BF764030	BF764030 IL2-CS004
5	17.8	84.8	425	3	AA195418	AA195418 zr36g10.s
c 6	17.8	84.8	612	237	AZ066229	AZ066229 RPCI-23-3
c 7	17.4	82.9	499	107	AU030401	AU030401 AU030401
8	17	81.0	624	168	BF700745	BF700745 602128542
c 9	16.8	80.0	437	11	AA778238	AA778238 zi97g08.s
c 10	16.8	80.0	440	107	AU070697	AU070697 AU070697
c 11	16.8	80.0	450	107	AU094227	AU094227 AU094227
c 12	16.8	80.0	484	151	BF598564	BF598564 sv19a09.y
13	16.8	80.0	520	155	BG557402	BG557402 EM1_43_E0
c 14	16.8	80.0	534	232	AQ737479	AQ737479 HS_3195_A
15	16.8	80.0	535	145	BF193631	BF193631 245281 MA
16	16.8	80.0	535	251	AZ911013	AZ911013 RPCI-24-1
17	16.8	80.0	561	145	BF194355	BF194355 246337 MA
18	16.8	80.0	583	244	AZ483876	AZ483876 1M0310P03
19	16.8	80.0	628	6	AA392024	AA392024 LD10961.5
20	16.8	80.0	629	6	AA392019	AA392019 LD10976.5
c 21	16.8	80.0	698	107	AU094228	AU094228 AU094228
22	16.8	80.0	713	14	AA951996	AA951996 LD28992.5
23	16.8	80.0	881	150	BF578486	BF578486 602092908
24	16.4	78.1	306	160	BB520416	BB520416 BB520416
25	16.4	78.1	346	120	AW764293	AW764293 ur74c08.x
26	16.4	78.1	522	142	BE942800	BE942800 EST422379
27	16.4	78.1	568	142	BE942802	BE942802 EST422381
c 28	16.4	78.1	683	240	AZ260166	AZ260166 RPCI-23-1
29	16.4	78.1	694	219	AG013064	AG013064 Homo sapi
c 30	16.4	78.1	697	175	BG282828	BG282828 602405927
31	16.4	78.1	1055	146	BF304360	BF304360 601887268
c 32	16.4	78.1	1071	220	CNS02E0E	AL193127 Tetraodon
33	16.2	77.1	219	149	BF490633	BF490633 AT27168.5
c 34	16.2	77.1	306	131	BB336747	BB336747 BB336747
35	16.2	77.1	317	245	AZ504982	AZ504982 1M0345O19
36	16.2	77.1	403	247	AZ647967	AZ647967 1M0514F17
37	16.2	77.1	412	118	AW590028	AW590028 hg30a04.x
38	16.2	77.1	416	172	BG019825	BG019825 dc68g01.x
39	16.2	77.1	419	246	AZ604463	AZ604463 1M0425O15
40	16.2	77.1	437	113	AW235361	AW235361 xm56a07.x
41	16.2	77.1	439	23	AI655576	AI655576 tt27h06.x
42	16.2	77.1	439	104	AI970357	AI970357 wq91c03.x
43	16.2	77.1	440	118	AW590489	AW590489 hg45e07.x
c 44	16.2	77.1	443	171	BF932393	BF932393 CM2-NT017
c 45	16.2	77.1	446	257	B98792	B98792 CIT-HSP-228

ALIGNMENTS

RESULT 1
T71061/c
LOCUS T71061 327 bp mRNA EST 01-MAR-1995
DEFINITION yc50a04.r1 Stratagene liver (#937224) Homo sapiens cDNA clone IMAGE:84078 5' similar to gb:M57899 UDP-GLUCURONOSYLTRANSFERASE 1A PRECURSOR, MICROSOMAL (HUMAN);, mRNA sequence.
ACCESSION T71061
VERSION T71061.1 GI:685582
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 327)
AUTHORS Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B., Chissoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W., Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N., Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Meg,J., Trevaskis,E., Underwood,K., Wohldmann,P., Waterston,R., Wilson,R. and Marra,M.
TITLE Generation and analysis of 280,000 human expressed sequence tags
JOURNAL Genome Res. 6 (9), 807-828 (1996)
MEDLINE 97044478
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Insert Size: 603
High qality sequence stops: 290 Source: IMAGE Consortium, LLNL This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 603 Std Error: 0.00
Seq primer: M13RP1
High quality sequence stop: 290.
FEATURES Location/Qualifiers
source 1..327
/organism="Homo sapiens"
/db_xref="GDB:501135"
/db_xref="taxon:9606"
/clone="IMAGE:84078"
/clone_lib="Stratagene liver (#937224)"
/sex="male"
/dev_stage="49 years old"
/lab_host="SOLR cells (kanamycin resistant)"
/note="Organ: liver; Vector: pBluescript SK; Site_1: EcoRI ; Site_2: XhoI; Cloned unidirectionally. Primer: Oligo dT. Hepatectomy from normal male caucasian. Average insert size: 1.1 kb; Uni-ZAP XR Vector; ~5' adaptor sequence: 5' GAATTCGGCACGAG 3' ~3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTTTT 3'"
BASE COUNT 71 a 74 c 92 g 83 t 7 others
ORIGIN

Query Match 100.0%; Score 21; DB 189; Length 327;
Best Local Similarity 100.0%; Pred. No. 2.2;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ccactgggatcaacagtatct 21
||||||||||||||||||||
Db 97 CCACTGGGATCAACAGTATCT 77

RESULT 2
AA195421
LOCUS AA195421 383 bp mRNA EST 06-AUG-1997
DEFINITION zr36h09.s1 Soares_NhHMPu_S1 Homo sapiens cDNA clone IMAGE:665537 3' similar to contains element MSR1 repetitive element ;, mRNA sequence.
ACCESSION AA195421
VERSION AA195421.1 GI:1785114
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 383)
AUTHORS Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.
TITLE WashU-Merck EST Project 1997
JOURNAL Unpublished (1997)
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810

Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1881 Std Error: 0.00
High quality sequence stop: 218.

FEATURES Location/Qualifiers
source 1. .383
/organism="Homo sapiens"
/db_xref="GDB:5427569"
/db_xref="taxon:9606"
/clone="IMAGE:665537"
/clone_lib="Soares_NhHMPu_S1"
/tissue_type="Pooled human melanocyte, fetal heart, and pregnant uterus"
/lab_host="DH10B"
/note="Organ: mixed (see below); Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; Equal amounts of plasmid DNA from three normalized libraries (melanocyte 2NbHM, pregnant uterus NbHPU, and fetal heart NbHH19W) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of I.M.A.G.E. clones 260232-265223, 340488-345479, and 484488-489479."

BASE COUNT 108 a 83 c 72 g 112 t 8 others
ORIGIN

Query Match 87.6%; Score 18.4; DB 3; Length 383;
Best Local Similarity 90.5%; Pred. No. 49;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ccactgggatcaacagtatct 21
||||||| |||| ||||||||
Db 180 CCACTGGCATCANCAGTATCT 200

RESULT 3
CNS04CB0/c
LOCUS CNS04CB0 1078 bp DNA GSS 21-MAY-2000
DEFINITION Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone 099M21 of library G from Tetraodon nigroviridis, genomic survey sequence.
ACCESSION AL284229
VERSION AL284229.1 GI:8022608
KEYWORDS GSS; genome survey sequence.
SOURCE Tetraodon nigroviridis.
ORGANISM Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodon.
REFERENCE 1 (bases 1 to 1078)
AUTHORS Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and Weissenbach,J.
TITLE Charaterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1078)
AUTHORS Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F., Saurin,W. and Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 1078)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/Tetraodon.

FEATURES Location/Qualifiers
source 1. .1078
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="099M21"
/clone_lib="G"
/note="Genoscope sequence ID : C0BG099AG11SP1-end : PUC-Ori"

BASE COUNT 343 a 192 c 230 g 312 t 1 others
ORIGIN

Query Match 87.6%; Score 18.4; DB 221; Length 1078;
Best Local Similarity 95.0%; Pred. No. 59;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ccactgggatcaacagtatc 20
||||||||||||||||| ||
Db 945 CCACTGGGATCAACAGTTTC 926

RESULT 4
BF764030/c
LOCUS BF764030 297 bp mRNA EST 12-JAN-2001
DEFINITION IL2-CS0048-301000-198-G10 CS0048 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF764030
VERSION BF764030.1 GI:12111930
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 297)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=IL2&t2=IL2-CS0048-301000-198-G10&t3=2000-10-30&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 58.

FEATURES Location/Qualifiers
source 1. .297
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="CS0048"
/dev_stage="Adult"
/note="Organ: colon_est; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

```
BASE COUNT       60 a     81 c     95 g     61 t
ORIGIN

  Query Match             84.8%;  Score 17.8;  DB 169;  Length 297;
  Best Local Similarity   90.5%;  Pred. No. 96;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 ccactgggatcaacagtatct 21
          ||||||||| | |||||||||
Db    128 CCACTGGGAGCGACAGTATCT 108


RESULT   5
AA195418
LOCUS       AA195418      425 bp    mRNA            EST       06-AUG-1997
DEFINITION  zr36g10.s1 Soares_NhHMPu_S1 Homo sapiens cDNA clone IMAGE:665538 3'
            similar to contains element MER5 repetitive element ;, mRNA
            sequence.
ACCESSION   AA195418
VERSION     AA195418.1  GI:1785111
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 425)
  AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J., Moore,B.
            , Schellenberg,K., Steptoe,M., Tan,F., Theising,B., White,Y., Wylie
            ,T., Waterston,R. and Wilson,R.
  TITLE     WashU-Merck EST Project 1997
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 1873    Std Error: 0.00
            High quality sequence stop: 337.
FEATURES             Location/Qualifiers
     source          1. .425
                     /organism="Homo sapiens"
                     /db_xref="GDB:5427570"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:665538"
                     /clone_lib="Soares_NhHMPu_S1"
                     /tissue_type="Pooled human melanocyte, fetal heart, and
                     pregnant uterus"
                     /lab_host="DH10B"
                     /note="Organ: mixed (see below); Vector: pT7T3D-Pac
                     (Pharmacia) with a modified polylinker; Site_1: Not I;
                     Site_2: Eco RI; Equal amounts of plasmid DNA from three
                     normalized libraries (melanocyte 2NbHM, pregnant uterus
                     NbHPU, and fetal heart NbHH19W) were mixed, and ss circles
                     were made in vitro. Following HAP purification, this DNA
                     was used as tracer in a subtractive hybridization
                     reaction. The driver was PCR-amplified cDNAs from pools of
                     5,000 clones made from the same 3 libraries. The pools
                     consisted of I.M.A.G.E. clones 260232-265223,
                     340488-345479, and 484488-489479."
BASE COUNT      129 a     85 c     80 g    129 t      2 others
ORIGIN

  Query Match             84.8%;  Score 17.8;  DB 3;  Length 425;
  Best Local Similarity   90.5%;  Pred. No. 1e+02;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 ccactgggatcaacagtatct 21
          ||||||||| | |||||||||
Db    180 CCACTGGCATCAGCAGTATCT 200


RESULT   6
AZ066229/c
LOCUS       AZ066229      612 bp    DNA             GSS       30-MAR-2000
DEFINITION  RPCI-23-392B4.TV RPCI-23 Mus musculus genomic clone RPCI-23-392B4,
            DNA sequence.
ACCESSION   AZ066229
VERSION     AZ066229.1  GI:7357481
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 612)
  AUTHORS   Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret
            ,B., Levins,M., Mcgann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
            and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-23
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-23. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
            or from Resea ch Genetics (info@resgen.com). BAC end page:
            http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 392  row: B  column: 4
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .612
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-23-392B4"
                     /clone_lib="RPCI-23"
                     /sex="Female"
                     /lab_host="DH10B"
                     /note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:
                     EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
                     brain genomic DNA was isolated and  partially digested
                     with a combination of EcoRI and EcoRI Methylase.   Size
                     selected DNA was cloned into the pBACe3.6 vector at the
                     EcoRI  sites.  The ligation products were transformed into
                     DH10B  electrocompetent cells (BRL Life Technologies)."
BASE COUNT      176 a    121 c    116 g    199 t
ORIGIN

  Query Match             84.8%;  Score 17.8;  DB 237;  Length 612;
  Best Local Similarity   90.5%;  Pred. No. 1.1e+02;
  Matches   19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 ccactgggatcaacagtatct 21
          || |||||||||||||| ||||
Db    540 CCCCTGGGATCAACAGAATCT 520


RESULT   7
AU030401/c
LOCUS       AU030401      499 bp    mRNA            EST       19-OCT-1998
DEFINITION  AU030401 Rice cDNA from immature leaf including apical meristem
            Oryza sativa cDNA clone E51014_42, mRNA sequence.
ACCESSION   AU030401
```

```
VERSION     AU030401.1  GI:3763649
KEYWORDS    EST.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 499)
  AUTHORS   Sasaki,T. and Yamamoto,K.
  TITLE     Rice cDNA from immature leaf including apical meristem
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Takuji Sasaki
            National Institute of Agrobiological Resources
            Rice Genome Research Program
            2-1-2 Kannondai,Tsukuba
            Ibaraki,
            Japan 305
            Tel: 0298-38-7441
            Fax: 0298-38-7468
            Email: tsasaki@abr.affrc.go.jp
            PROJECT ='RGP'.
FEATURES             Location/Qualifiers
     source          1. .499
                     /organism="Oryza sativa"
                     /strain="Nipponbare"
                     /db_xref="taxon:4530"
                     /clone="E51014_4Z"
                     /clone_lib="Rice cDNA from immature leaf including apical
                     meristem"
                     /dev_stage="immature"
                     /note="Organ: leaf; immature leaf including apical
                     meristem (under long day condition)"
BASE COUNT      132 a    112 c    133 g    115 t      7 others
ORIGIN
```

```
  Query Match           82.9%;  Score 17.4;  DB 107;  Length 499;
  Best Local Similarity 90.0%;  Pred. No. 1.7e+02;
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 ccactgggatcaacagtatc 20
           |||||| |||||||||| ||
Db      78 CCACTGNGATCAACAGTTTC 59
```

```
RESULT   8
BF700745
LOCUS       BF700745      624 bp    mRNA            EST       22-DEC-2000
DEFINITION  602128542F1 NIH_MGC_56 Homo sapiens cDNA clone IMAGE:4285378 5',
            mRNA sequence.
ACCESSION   BF700745
VERSION     BF700745.1  GI:11986153
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 624)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: CLONETECH Laboratories, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1120  row: c  column: 11
            High quality sequence stop: 623.
FEATURES             Location/Qualifiers
     source          1. .624
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4285378"
                     /clone_lib="NIH_MGC_56"
                     /tissue_type="primitive neuroectoderm"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: brain; Vector: pDNR-LIB (Clontech); Site_1:
                     SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc);
                     Double-stranded cDNA was prepared from cell line RNA.  5'
                     and 3' adaptors were used in cloning as follows: 5'
                     adaptor sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor
                     sequence:  5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3'
                     (where B = A, C, or G and N = A, C,  G, or T).   Average
                     insert size 1.65 kb (range 0.9-4.0 kb).  15/15 colonies
                     contained  inserts by PCR. This library was enriched for
                     full-length clones and was constructed by Clontech
                     Laboratories (Palo Alto, CA)."
BASE COUNT      125 a    175 c    182 g    142 t
ORIGIN
```

```
  Query Match           81.0%;  Score 17;  DB 168;  Length 624;
  Best Local Similarity 100.0%;  Pred. No. 2.8e+02;
  Matches   17;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ccactgggatcaacagt 17
           |||||||||||||||||
Db     351 CCACTGGGATCAACAGT 367
```

```
RESULT   9
AA778238/c
LOCUS       AA778238      437 bp    mRNA            EST       05-FEB-1998
DEFINITION  zi97g08.s1 Soares_fetal_liver_spleen_1NFLS_S1 Homo sapiens cDNA
            clone IMAGE:448766 3', mRNA sequence.
ACCESSION   AA778238
VERSION     AA778238.1  GI:2836953
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 437)
  AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin
            ,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,
            White,Y., Wylie,T., Waterston,R. and Wilson,R.
  TITLE     WashU-NCI human EST Project
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 399.
FEATURES             Location/Qualifiers
     source          1. .437
                     /organism="Homo sapiens"
                     /db_xref="GDB:1353023"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:448766"
                     /clone_lib="Soares_fetal_liver_spleen_1NFLS_S1"
                     /sex="male"
                     /dev_stage="20 week-post conception fetus"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
                     with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
                     This is a subtracted version of the original Soares fetal
```

liver spleen 1NFLS library. 1st strand cDNA was primed with a Pac I - oligo(dT) primer [5' AACTGGAAGAATTAATTAAAGATCTTTTTTTTTTTTTTTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT 116 a 104 c 99 g 118 t
ORIGIN

Query Match 80.0%; Score 16.8; DB 11; Length 437;
Best Local Similarity 90.0%; Pred. No. 3.4e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 cactgggatcaacagtatct 21
||||||||||||||| ||| |
Db 315 CACTGGGATCAACAATATTT 296

RESULT 10
AU070697/c
LOCUS AU070697 440 bp mRNA EST 10-JUN-1999
DEFINITION AU070697 Rice cDNA from young root Oryza sativa cDNA clone R10115_1A, mRNA sequence.
ACCESSION AU070697
VERSION AU070697.1 GI:5038587
KEYWORDS EST.
SOURCE Oryza sativa.
ORGANISM Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.
REFERENCE 1 (bases 1 to 440)
AUTHORS Yamamoto,K. and Sasaki,T.
TITLE Rice cDNA from young root
JOURNAL Unpublished (1999)
COMMENT Contact: Takuji Sasaki
National Institute of Agrobiological Resources
Rice Genome Research Program
2-1-2 Kannondai,Tsukuba
Ibaraki,
Japan 305
Tel: 0298-38-7441
Fax: 0298-38-7468
Email: tsasaki@abr.affrc.go.jp
PROJECT ='RGP'.
FEATURES Location/Qualifiers
source 1. .440
/organism="Oryza sativa"
/strain="Nipponbare"
/db_xref="taxon:4530"
/clone="R10115_1A"
/clone_lib="Rice cDNA from young root"
/tissue_type="young root"
BASE COUNT 126 a 88 c 94 g 132 t
ORIGIN

Query Match 80.0%; Score 16.8; DB 107; Length 440;
Best Local Similarity 90.0%; Pred. No. 3.4e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 cactgggatcaacagtatct 21
|| ||||||||| ||||||||
Db 29 CAGTGGGATCAGCAGTATCT 10

RESULT 11
AU094227/c
LOCUS AU094227 450 bp mRNA EST 30-JUN-2000
DEFINITION AU094227 Rice panicle at flowering stage Oryza sativa cDNA clone E3685, mRNA sequence.
ACCESSION AU094227
VERSION AU094227.1 GI:8856909
KEYWORDS EST.
SOURCE Oryza sativa.
ORGANISM Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.
REFERENCE 1 (bases 1 to 450)
AUTHORS Sasaki,T. and Yamamoto,K.
TITLE Rice cDNA from panicle at flowering stage (2000)
JOURNAL Unpublished (2000)
COMMENT Contact: Takuji Sasaki
National Institute of Agrobiological Resources
Rice Genome Research Program
2-1-2 Kannondai,Tsukuba
Ibaraki,
Japan 305
Tel: 0298-38-7441
Fax: 0298-38-7468
Email: tsasaki@abr.affrc.go.jp
PROJECT ='RGP'.
E3685_3A.
FEATURES Location/Qualifiers
source 1. .450
/organism="Oryza sativa"
/strain="Nipponbare"
/db_xref="taxon:4530"
/clone="E3685"
/clone_lib="Rice panicle at flowering stage"
/dev_stage="flowering stage"
/note="Organ: panicle; Rice cDNA from panicle at flowering stage"
BASE COUNT 117 a 91 c 121 g 120 t 1 others
ORIGIN

Query Match 80.0%; Score 16.8; DB 107; Length 450;
Best Local Similarity 90.0%; Pred. No. 3.4e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 cactgggatcaacagtatct 21
|| ||||||||| ||||||||
Db 254 CAGTGGGATCAGCAGTATCT 235

RESULT 12
BF598564/c
LOCUS BF598564 484 bp mRNA EST 12-DEC-2000
DEFINITION sv19a09.y1 Gm-c1057 Glycine max cDNA clone GENOME SYSTEMS CLONE ID: Gm-c1057-17 5' similar to SW:PLAS_LYCES P17340 PLASTOCYANIN PRECURSOR. [1] ;, mRNA sequence.
ACCESSION BF598564
VERSION BF598564.1 GI:11690888
KEYWORDS EST.
SOURCE soybean.
ORGANISM Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Glycine.
REFERENCE 1 (bases 1 to 484)
AUTHORS Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.
TITLE Public Soybean EST Project
JOURNAL Unpublished (1999)
COMMENT Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134 For further information call: (800) 430-0030 or (314) 427-3222 FAX:(888) 919-3324 or (314) 427-3324 or contact: clones@genomesystems.com or info@genomesystems.com web site: www.genomesystems.com
Trace considered overall poor quality
High quality sequence stop: 1.

FEATURES Location/Qualifiers
source 1. .484
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-c1057-17"
/clone_lib="Gm-c1057"
/tissue_type="Degenerating cotyledons, 2 week old seedling"
/lab_host="DH10B"
/note="Vector: pBluescript II SK+; Site_1: EcoRI; Site_2: XhoI; The cDNA library was constructed from mRNA isolated from degenerating cotyledons of 2 week old seedlings from PI468916. Complementary DNA was synthesized from mRNA using a primer consisting of a poly(dT) sequence with a XhoI restriction site. EcoRI adapters were ligated to the blunt-ended cDNA fragments followed by XhoI digestion. The cDNA fragments were directionally cloned into the EcoRI-XhoI restriction site of the pBluescript vector. The ligated cDNA fragments were transformed into DH10B host cells (GibcoBRL). This library was constructed in the laboratory of Dr. Randy Shoemaker."

BASE COUNT 117 a 133 c 121 g 113 t
ORIGIN

Query Match 80.0%; Score 16.8; DB 151; Length 484;
Best Local Similarity 90.0%; Pred. No. 3.4e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy      1 ccactgggatcaacagtatc 20
          ||| ||| ||||||||||||
Db    435 CCATTGGCATCAACAGTATC 416
```

RESULT 13
BG557402
LOCUS BG557402 520 bp mRNA EST 10-APR-2001
DEFINITION EM1_43_E08.b1_A002 Embryo 1 (EM1) Sorghum bicolor cDNA, mRNA sequence.
ACCESSION BG557402
VERSION BG557402.1 GI:13586400
KEYWORDS EST.
SOURCE sorghum.
ORGANISM Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE 1 (bases 1 to 520)
AUTHORS Reid,S.P., Cordonnier-Pratt,M.-M., Gingle,A. and Pratt,L.H.
TITLE An EST database from Sorghum: developing embryos
JOURNAL Unpublished (2000)
COMMENT Contact: Cordonnier-Pratt MM
Department of Botany
The University of Georgia
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 542 1805
Email: mmpratt@uga.edu
Sequences have been trimmed to exclude PolyA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20.
Seq primer: JEN REV
High quality sequence stop: 510
POLYA=No.

FEATURES Location/Qualifiers
source 1. .520
/organism="Sorghum bicolor"
/db_xref="taxon:4558"
/clone_lib="Embryo 1 (EM1)"
/note="Organ: Embryos germinated for 24 hr; Vector: pBluescript II from Lambda Zap II; Site_1: XhoI; Site_2: EcoRI; The library was made from poly-A RNA in the cloning vector lambda ZAP II. Clones to be sequenced were prepared by mass excision."

BASE COUNT 121 a 164 c 112 g 122 t 1 others
ORIGIN

Query Match 80.0%; Score 16.8; DB 155; Length 520;
Best Local Similarity 90.0%; Pred. No. 3.5e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy      2 cactgggatcaacagtatct 21
          ||| ||||||||||||||||
Db      2 CACGAGGATCAACAGTATCT 21
```

RESULT 14
AQ737479/c
LOCUS AQ737479 534 bp DNA GSS 16-JUL-1999
DEFINITION HS_3195_A2_B02_T7C CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=3195 Col=4 Row=C, DNA sequence.
ACCESSION AQ737479
VERSION AQ737479.1 GI:5515001
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 534)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones may be purchased from Research Genetics (info@resgen.com).
BAC end Web Server: http://www.htsc.washington.edu
Plate: 3195 row: C column: 4
Seq primer: T7
Class: BAC ends
High quality sequence stop: 534.

FEATURES Location/Qualifiers
source 1. .534
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=3195 Col=4 Row=C"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in E-Coli DH10B"

BASE COUNT 140 a 120 c 121 g 146 t 7 others
ORIGIN

Query Match 80.0%; Score 16.8; DB 232; Length 534;
Best Local Similarity 90.0%; Pred. No. 3.5e+02;

```
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       2 cactgggatcaacagtatct 21
           |||||||| |||||||| ||
Db     210 CACTGGGAGCAACAGTACCT 191


RESULT  15
BF193631
LOCUS       BF193631      535 bp    mRNA            EST       02-NOV-2000
DEFINITION  245281 MARC 2PIG Sus scrofa cDNA 5', mRNA sequence.
ACCESSION   BF193631
VERSION     BF193631.1  GI:11077000
KEYWORDS    EST.
SOURCE      pig.
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 535)
  AUTHORS   Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,
            Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W.
            and Keele,J.W.
  TITLE     Design and use of two pooled tissue normalized cDNA libraries for
            EST discovery in swine
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 74  row: K  column: 12
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES             Location/Qualifiers
     source          1. .535
                     /organism="Sus scrofa"
                     /db_xref="taxon:9823"
                     /clone_lib="MARC 2PIG"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
                     Library made from pooled tissue from testis, ovary,
                     endometrium, hypothalamus, pituitary, and placenta."
BASE COUNT      163 a    136 c    118 g    118 t
ORIGIN


  Query Match           80.0%;  Score 16.8;  DB 145;  Length 535;
  Best Local Similarity  90.0%;  Pred. No. 3.5e+02;
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 ccactgggatcaacagtatc 20
           |||||||||||||||| | |
Db     514 CCACTGGGATCAACAGCAAC 533


Search completed: July 25, 2001, 04:54:51
Job time: 10424 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: July 25, 2001, 05:16:58 ; Search time 1290.33 Seconds
(without alignments)
239.749 Million cell updates/sec

Title: US-09-142-095-3
Perfect score: 20
Sequence: 1 gtcacgtgacacagtcaaac 20

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues

Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_ba3:*
4: gb_in1:*
5: gb_in2:*
6: gb_in3:*
7: gb_om:*
8: gb_ov:*
9: gb_pat1:*
10: gb_pat2:*
11: gb_ph:*
12: gb_pl1:*
13: gb_pl2:*
14: gb_pl3:*
15: gb_pl4:*
16: em_ba1:*
17: em_ba2:*
18: em_fun:*
19: em_htgo_hum:*
20: em_htgo_inv:*
21: em_htgo_rod:*
22: em_htg_hum1:*
23: em_htg_hum2:*
24: em_htg_hum3:*
25: em_htg_hum4:*
26: em_htg_hum5:*
27: em_htg_hum6:*
28: em_htg_hum7:*
29: em_htg_hum8:*
30: em_htg_inv1:*
31: em_htg_inv2:*
32: em_htg_other:*
33: em_htg_rod:*
34: em_hum1:*
35: em_hum2:*
36: em_hum3:*
37: em_hum4:*
38: em_hum5:*
39: em_hum6:*
40: em_hum7:*
41: em_in:*
42: em_om:*
43: em_or:*
44: em_ov:*
45: em_pat:*
46: em_ph:*
47: em_pl:*
48: em_ro:*
49: em_sts:*
50: em_sy:*
51: em_un:*
52: em_vi:*
53: gb_sts1:*
54: gb_sts2:*
55: gb_sts3:*
56: gb_sy:*
57: gb_un:*
58: gb_vi1:*
59: gb_vi2:*
60: gb_htg1:*
61: gb_htg2:*
62: gb_htg3:*
63: gb_htg4:*
64: gb_htg5:*
65: gb_htg6:*
66: gb_htg7:*
67: gb_htg8:*
68: gb_htg9:*
69: gb_htg10:*
70: gb_htg11:*
71: gb_htg12:*
72: gb_htg13:*
73: gb_htg14:*
74: gb_htg15:*
75: gb_htg16:*
76: gb_htg17:*
77: gb_htg18:*
78: gb_htg19:*
79: gb_htg20:*
80: gb_htg21:*
81: gb_htg22:*
82: gb_htg23:*
83: gb_htg24:*
84: gb_htg25:*
85: gb_pr1:*
86: gb_pr2:*
87: gb_pr3:*
88: gb_pr4:*
89: gb_pr5:*
90: gb_pr6:*
91: gb_pr7:*
92: gb_pr8:*
93: gb_pr9:*
94: gb_ro1:*
95: gb_ro2:*
96: gb_in4:*
97: gb_pr10:*
98: em_ba3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		20	100.0	20	9	A65502	A65502 Sequence 3
2		20	100.0	531	89	AF352795	AF352795 Homo sapi
3		20	100.0	541	89	AF180372	AF180372 Homo sapi
4		20	100.0	620	9	A65504	A65504 Sequence 5
5		20	100.0	3341	91	D87674	D87674 Homo sapien
6	C	20	100.0	68770	86	AC006985	AC006985 Homo sapi
7	C	20	100.0	176619	70	AC026497	AC026497 Homo sapi
8		20	100.0	198872	89	AF297093	AF297093 Homo sapi

	9	18.4	92.0	200	89	AF135466
	10	18.4	92.0	208	89	AF135467
	11	18.4	92.0	234	89	AF135462
	12	18.4	92.0	234	89	AF135463
	13	17.4	87.0	4201	94	MUSPC326A
c	14	17.4	87.0	99123	12	AF069716
	15	16.8	84.0	207	89	AF135470
	16	16.8	84.0	208	89	AF135464
	17	16.8	84.0	208	89	AF135468
	18	16.8	84.0	208	89	AF135469
	19	16.8	84.0	208	89	AF135471
c	20	16.8	84.0	917	53	CNS06YPM
	21	16.8	84.0	3459	8	AF239838
c	22	16.8	84.0	145481	90	AL356862
	23	16.8	84.0	172697	71	AC027673
c	24	16.8	84.0	175793	86	AC005920
	25	16.8	84.0	178004	80	AL357623
	26	16.8	84.0	186323	87	AC009405
c	27	16.8	84.0	200085	64	AC016818
	28	16.8	84.0	251124	97	HUAE000660
c	29	16.4	82.0	22	9	A67543
c	30	16.4	82.0	319	54	G45143
c	31	16.4	82.0	5961	54	G31676
c	32	16.4	82.0	5961	93	HSBSF2
c	33	16.4	82.0	90736	92	HS299D3
	34	16.4	82.0	112031	90	AL445487
	35	16.4	82.0	150351	94	AC074047
c	36	16.4	82.0	163732	74	AC073072
c	37	16.4	82.0	199949	68	AC023777
	38	16.4	82.0	208393	77	AC087870
c	39	16.4	82.0	211774	75	AC074048
	40	16.4	82.0	274365	62	AC012295
c	41	16	80.0	2573	94	AF190798
	42	16	80.0	106553	77	AC084845
c	43	16	80.0	192493	68	AC024731
	44	16	80.0	309401	64	AC016584
	45	15.8	79.0	188	7	AB010592

AF135466 Pongo pyg
AF135467 Trachypit
AF135462 Pan panis
AF135463 Pan trogl
M95564 Mus musculu
AF069716 Arabidops
AF135470 Cebus ape
AF135464 Gorilla g
AF135468 Papio cyn
AF135469 Colobus g
AF135471 Saimiri b
AL421328 T7 end of
AF239838 Gallus ga
AL356862 Human DNA
AC027673 Homo sapi
AC005920 Homo sapi
AL357623 Homo sapi
AC009405 Homo sapi
AC016818 Homo sapi
AE000660 Homo sapi
A67543 Sequence 3
G45143 Z1989_1 Zeb
G31676 sWSS1282 Er
Y00081 Human (BSF-
Z84468 Human DNA s
AL445487 Human DNA
AC074047 Mus muscu
AC073072 Homo sapi
AC023777 Homo sapi
AC087870 Mus muscu
AC074048 Mus muscu
AC012295 Mus muscu
AF190798 Rattus no
AC084845 Homo sapi
AC024731 Homo sapi
AC016584 Homo sapi
AB010592 Globiceph

ALIGNMENTS

```
RESULT   1
A65502
LOCUS       A65502          20 bp    DNA              PAT       29-MAR-1999
DEFINITION  Sequence 3 from Patent WO9732042.
ACCESSION   A65502
VERSION     A65502.1  GI:4531237
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 20)
  AUTHORS   Burchell,B.
  TITLE     DRUG TRIAL ASSAY SYSTEM
  JOURNAL   Patent: WO 9732042-A 3 04-SEP-1997;
            UNIV DUNDEE (GB)
COMMENT     Other publication AU 2224197 19970916.
FEATURES             Location/Qualifiers
     source          1. .20
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT        7 a      6 c      4 g      3 t
ORIGIN
```

```
  Query Match           100.0%;  Score 20;  DB 9;  Length 20;
  Best Local Similarity  100.0%;  Pred. No. 0.65;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gtcacgtgacacagtcaaac 20
           ||||||||||||||||||||
Db       1 GTCACGTGACACAGTCAAAC 20
```

```
RESULT   2
AF352795
LOCUS       AF352795       531 bp    DNA              PRI       10-APR-2001
DEFINITION  Homo sapiens bilirubin UDP-glucronosyltrasferase 1-1 (UGT1A1) gene,
            UGT1A1*1 allele, partial cds.
ACCESSION   AF352795
VERSION     AF352795.1  GI:13569708
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 531)
  AUTHORS   McKie,K., Kutlar,F., Glendenning,M. and Kutlar,A.
  TITLE     7(TA) repeat polymorphism of the TATA box of human bilirubin
            UDP-glucuronosyltransferase 1-1(UGT1A1*1) gene in a patient with
            sickle cell anemia + high bilirubinemia
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 531)
  AUTHORS   McKie,K., Kutlar,F., Glendenning,M. and Kutlar,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-FEB-2001) Medicine/Hemoglobin DNA Laboratory, Medical
            College of Georgia, 15th St., AC-1000, Augusta, GA 30912, USA
FEATURES             Location/Qualifiers
     source          1. .531
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /map="2q37"
                     /cell_type="white blood cell"
                     /tissue_type="blood"
     gene            <1. .>531
                     /gene="UGT1A1"
                     /note="UGT1"
                     /allele="UGT1A1*1"
     promoter        <1. .313
                     /gene="UGT1A1"
     TATA_signal     279. .295
                     /gene="UGT1A1"
     repeat_region   279. .>292
                     /note="repeat polymorphism compared to UGT1A1 sequence
                     presented in GenBank Accession Number AF180372; contains 7
                     ta repeats frequently detected in patients with high
                     bilirubinemia"
                     /rpt_type=tandem
                     /rpt_unit=ta
     mRNA            314. .>531
                     /gene="UGT1A1"
                     /product="bilirubin UDP-glucronosyltrasferase 1-1"
     CDS             334. .>531
                     /gene="UGT1A1"
                     /codon_start=1
                     /product="bilirubin UDP-glucronosyltrasferase 1-1"
                     /protein_id="AAK31204.1"
                     /db_xref="GI:13569709"
                     /translation="MAVESQGGRPLVLGLLLCVLGPVVSHAGKILLIPVDGSHWLSML
                     GAIQQLQQRGHEIVVLAPDASL"
BASE COUNT      120 a    121 c    137 g    153 t
ORIGIN
```

```
  Query Match           100.0%;  Score 20;  DB 89;  Length 531;
  Best Local Similarity  100.0%;  Pred. No. 0.58;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gtcacgtgacacagtcaaac 20
           ||||||||||||||||||||
Db     228 GTCACGTGACACAGTCAAAC 247
```

RESULT 3

```
AF180372
LOCUS       AF180372      541 bp    DNA             PRI       05-OCT-1999
DEFINITION  Homo sapiens bilirubin UDP-glucuronosyltransferase 1-1 (UGT1) gene,
            UGT1*1 allele, partial cds.
ACCESSION   AF180372
VERSION     AF180372.1  GI:6010649
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 541)
  AUTHORS   Kutlar,F., Sromek,E., Leithner,C., Nechtman,J. and Kutlar,A.
  TITLE     Detection of the TATA box polymorphism of the human bilirubin
            UDP-glucuronosyltransferase 1-1 gene (UGT1*1) in a patient with
            sickle cell anemia
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 541)
  AUTHORS   Kutlar,F., Sromek,E., Leithner,C., Nechtman,J. and Kutlar,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-AUG-1999) Medicine, Hematology/Oncology-Sickle Cell
            Center, Medical College of Georgia, 15th Street, AC-1000, Augusta,
            GA 30912, USA
FEATURES             Location/Qualifiers
     source          1. .541
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /map="2q37"
                     /cell_type="white blood cells"
                     /tissue_type="whole blood"
     gene            1. .>541
                     /gene="UGT1"
                     /note="GNT1; UGT1A1"
                     /allele="UGT1*1"
     promoter        1. .313
                     /gene="UGT1"
     variation       226
                     /gene="UGT1"
                     /replace="t"
     repeat_region   279. .288
                     /note="polymorphic region"
                     /rpt_type=tandem
                     /rpt_unit=ta
     TATA_signal     279. .291
                     /gene="UGT1"
                     /note="Ritter,J.K., et al., 1992, J. Biol. Chem.,
                     267:3257-3261"
     mRNA            314. .>541
                     /gene="UGT1"
                     /product="bilirubin UDP-glucuronosyltransferase 1-1"
     CDS             330. .>541
                     /gene="UGT1"
                     /note="UDP glycosyltransferase 1"
                     /codon_start=1
                     /product="bilirubin UDP-glucuronosyltransferase 1-1"
                     /protein_id="AAF01205.1"
                     /db_xref="GI:6010650"
                     /translation="MAVESQGGRPLVLGLLLCVLGPVVSHAGKILLIPVDGSHWLSML
                     GAIQQLQQRGHEIVVLAPDASLYIRDG"
BASE COUNT      123 a    124 c    141 g    153 t
ORIGIN


  Query Match           100.0%;  Score 20;  DB 89;  Length 541;
  Best Local Similarity  100.0%;  Pred. No. 0.58;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gtcacgtgacacagtcaaac 20
           ||||||||||||||||||||
Db     228 GTCACGTGACACAGTCAAAC 247
```

```
RESULT   4
A65504
LOCUS       A65504        620 bp    DNA             PAT       29-MAR-1999
DEFINITION  Sequence 5 from Patent WO9732042.
ACCESSION   A65504
VERSION     A65504.1  GI:4531239
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 620)
  AUTHORS   Burchell,B.
  TITLE     DRUG TRIAL ASSAY SYSTEM
  JOURNAL   Patent: WO 9732042-A 5 04-SEP-1997;
            UNIV DUNDEE (GB)
COMMENT     Other publication AU 2224197 19970916.
FEATURES             Location/Qualifiers
     source          1. .620
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      157 a    127 c    151 g    185 t
ORIGIN


  Query Match           100.0%;  Score 20;  DB 9;  Length 620;
  Best Local Similarity  100.0%;  Pred. No. 0.58;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gtcacgtgacacagtcaaac 20
           ||||||||||||||||||||
Db     508 GTCACGTGACACAGTCAAAC 527


RESULT   5
D87674
LOCUS       D87674       3341 bp    DNA             PRI       14-APR-2000
DEFINITION  Homo sapiens gene for bilirubin UDP-glucuronosyltransferase 1,
            promoter region and partial cds.
ACCESSION   D87674
VERSION     D87674.1  GI:3059176
KEYWORDS    bilirubin UDP-glucuronosyltransferase 1.
SOURCE      Homo sapiens DNA.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Ueyama,H., Koiwai,O., Soeda,Y., Sato,H., Satoh,Y., Ohkubo,I. and
            Doida,Y.
  TITLE     Analysis of the promoter of human bilirubin
            UDP-glucuronosyltransferase gene (UGT1*1) in relevance to Gilbert's
            syndrome
  JOURNAL   Hepatol. Res. 9, 152-163 (1997)
REFERENCE   2  (bases 1 to 3341)
  AUTHORS   Ueyama,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-SEP-1996) to the DDBJ/EMBL/GenBank databases. Hisao
            Ueyama, Shiga University of Medical Science, Department of Medical
            Biochemistry; Seta, Otsu, Shiga 520-21, Japan (Tel:077-548-2162,
            Fax:077-548-2164)
COMMENT     Sequence updated (08-Jan-1997) by: Hisao Ueyama.
FEATURES             Location/Qualifiers
     source          1. .3341
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="2q37"
     misc_feature    2538. .2544
                     /note="XRE"
     misc_feature    2606. .2610
                     /note="XRE"
     misc_feature    3088. .3097
                     /note="E-box"
     misc_feature    3101. .3113
                     /note="HNF-1 site"
```

```
     CAAT_signal     3125. .3129
     TATA_signal     3149. .3153
     exon            3177. .3341
                     /number=1
     gene            3192. .3341
                     /gene="UGT1*1"
     CDS             3192. .>3341
                     /gene="UGT1*1"
                     /codon_start=1
                     /product="bilirubin UDP-glucuronosyltransferase 1"
                     /protein_id="BAA25600.1"
                     /db_xref="GI:3059177"
                     /translation="MAVESQGGRPLVLGLLLCVLGPVVSHAGKILLIPVDGSHWLSML
                     GAIQQL"
BASE COUNT      893 a    695 c    803 g    950 t
ORIGIN


  Query Match           100.0%;  Score 20;  DB 91;  Length 3341;
  Best Local Similarity  100.0%;  Pred. No. 0.55;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gtcacgtgacacagtcaaac 20
          ||||||||||||||||||||
Db   3088 GTCACGTGACACAGTCAAAC 3107


RESULT   6
AC006985/c
LOCUS       AC006985    68770 bp    DNA             PRI       21-DEC-1999
DEFINITION  Homo sapiens BAC clone RP11-154L24 from 2, complete sequence.
ACCESSION   AC006985
VERSION     AC006985.2  GI:5732165
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 68770)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
  MEDLINE   99063792
REFERENCE   2  (bases 1 to 68770)
  AUTHORS   Gattung,S., Stoneking,T. and Davidson,T.
  TITLE     The sequence of Homo sapiens BAC clone RP11-154L24
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 68770)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-MAR-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 68770)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-AUG-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   5  (bases 1 to 68770)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-OCT-1999) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   6  (bases 1 to 68770)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-DEC-1999) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Aug 13, 1999 this sequence version replaced gi:4337256.
            ---------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu/gsc
                Contact: sapiens@watson.wustl.edu
            ---------- Summary Statistics
                Center project name: H_NH0154L24
            ----------

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            Mapping information for this clone was provided by Dr. John D.
            McPherson, Department of Genetics, Washington University, St. Louis
            MO.  For additional information about the map position of this
            sequence, see http://genome.wustl.edu/gsc

            SOURCE INFORMATION:
            The RPCI-11 human BAC library was made from the blood of one male
            donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
            Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
            approach for construction of bacterial artificial chromosome
            libraries.  Genomics 51:1-8.  The clone may be obtained either from
            Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
            and coworkers at the Roswell Park Cancer Institute
            (http://bacpac.med.buffalo.edu)
            VECTOR:  pBACe3.6
            NEIGHBORING SEQUENCE INFORMATION:
            The clone sequenced to the left is RP11-332L11, 200 bp overlap.
            Actual start of this clone is at base position 86134 of
            RP11-332L11; actual end is at base position 68770 of RP11-154L24.

            The clone RP11-154L24 contains a tandem repeat from.base positions
            38234 to 39039, this region contains some low quality data. The
            assembly is consistent with the restriction digest information.
FEATURES                Location/Qualifiers
     source             1. .68770
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /chromosome="2"
                        /map="2"
                        /clone="RP11-154L24"
                        /clone_lib="RPCI-11"
     repeat_region      3. .192
                        /rpt_family="L1"
     repeat_region      221. .528
                        /rpt_family="Alu"
     repeat_region      529. .1634
                        /rpt_family="L1"
     repeat_region      1658. .1878
                        /rpt_family="MIR"
     repeat_region      2079. .2271
                        /rpt_family="MER1_type"
     repeat_region      2276. .2336
                        /rpt_family="MER1_type"
     repeat_region      4683. .4806
                        /rpt_family="MIR"
     repeat_region      5131. .5389
                        /rpt_family="MaLR"
     repeat_region      8738. .8871
                        /rpt_family="MIR"
     repeat_region      9092. .9371
                        /rpt_family="AcHobo"
     repeat_region      9967. .10582
                        /rpt_family="L1"
```

```
repeat_region   10576. .11200
                /rpt_family="L1"
repeat_region   11196. .11469
                /rpt_family="L1"
repeat_region   11540. .11654
                /rpt_family="L1"
repeat_region   12108. .12523
                /rpt_family="MaLR"
repeat_region   13113. .13149
                /rpt_family="MER1_type"
repeat_region   13150. .13171
                /rpt_family="(CAG)n"
repeat_region   13172. .13286
                /rpt_family="MER1_type"
repeat_region   13338. .13497
                /rpt_family="MER1_type"
repeat_region   14727. .14848
                /rpt_family="MIR"
repeat_region   15890. .16244
                /rpt_family="MaLR"
repeat_region   16380. .16723
                /rpt_family="L1"
repeat_region   17483. .17560
                /rpt_family="L2"
repeat_region   17655. .17747
                /rpt_family="L2"
repeat_region   18220. .18337
                /rpt_family="MER1_type"
repeat_region.  18770. .18816
                /rpt_family="(CA)n"
repeat_region   19103. .19121
                /rpt_family="AT_rich"
repeat_region   19122. .19636
                /rpt_family="L1"
repeat_region   19637. .19662
                /rpt_family="AT_rich"
repeat_region   19825. .20014
                /rpt_family="L1"
repeat_region   20034. .20098
                /rpt_family="MIR"
repeat_region   20254. .21067
                /rpt_family="L1"
repeat_region   21141. .21182
                /rpt_family="A-rich"
repeat_region   24649. .25018
                /rpt_family="Retroviral"
repeat_region   25258. .25447
                /rpt_family="MIR"
repeat_region   25777. .25824
                /rpt_family="(CA)n"
repeat_region   28020. .28178
                /rpt_family="MIR"
repeat_region   28178. .28222
                /rpt_family="L2"
repeat_region   28503. .28558
                /rpt_family="L1"
repeat_region   28558. .28741
                /rpt_family="L1"
repeat_region   28970. .29064
                /rpt_family="L1"
repeat_region   29276. .29434
                /rpt_family="MER1_type"
repeat_region   29554. .29712
                /rpt_family="MER1_type?"
repeat_region   30867. .31015
                /rpt_family="MER1_type"
repeat_region   31268. .31682
                /rpt_family="MaLR"
repeat_region   32195. .32390
                /rpt_family="MIR"
repeat_region   33607. .33760
                /rpt_family="L2"
repeat_region   33835. .34127
                /rpt_family="Alu"
repeat_region   34167. .34283
                /rpt_family="GA-rich"
repeat_region   34806. .34976
                /rpt_family="L2"
repeat_region   35247. .35719
                /rpt_family="L2"
repeat_region   35720. .35762
                /rpt_family="GA-rich"
repeat_region   35763. .35909

  Query Match            100.0%;  Score 20;  DB 86;  Length 68770;
  Best Local Similarity  100.0%;  Pred. No. 0.5;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gtcacgtgacacagtcaaac 20
          ||||||||||||||||||||
Db  66259 GTCACGTGACACAGTCAAAC 66240


RESULT   7
AC026497/c
LOCUS       AC026497   176619 bp    DNA             HTG       22-MAR-2000
DEFINITION  Homo sapiens chromosome 11 clone RP11-689A10 map 11, *** SEQUENCING
            IN PROGRESS ***, 43 unordered pieces.
ACCESSION   AC026497
VERSION     AC026497.1  GI:7283186
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 176619)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 11, clone RP11-689A10
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 176619)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-MAR-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L6277
                Center clone name: 689_A_10
            --------------
```

```
* NOTE: This is a 'working draft' sequence. It currently
* consists of 43 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*       1     1333: contig of 1333 bp in length
*    1334  1433: gap of      100 bp
*    1434     2787: contig of 1354 bp in length
*    2788  2887: gap of      100 bp
*    2888     3998: contig of 1111 bp in length
*    3999  4098: gap of      100 bp
*    4099     5777: contig of 1679 bp in length
*    5778  5877: gap of      100 bp
*    5878     7685: contig of 1808 bp in length
*    7686  7785: gap of      100 bp
*    7786     8176: contig of 391 bp in length
*    8177  8276: gap of      100 bp
*    8277     9926: contig of 1650 bp in length
*    9927 10026: gap of      100 bp
*   10027    11499: contig of 1473 bp in length
*   11500 11599: gap of      100 bp
*   11600    13616: contig of 2017 bp in length
*   13617 13716: gap of      100 bp
*   13717    16073: contig of 2357 bp in length
*   16074 16173: gap of      100 bp
*   16174    18510: contig of 2337 bp in length
*   18511 18610: gap of      100 bp
*   18611    20495: contig of 1885 bp in length
*   20496 20595: gap of      100 bp
*   20596    23548: contig of 2953 bp in length
*   23549 23648: gap of      100 bp
*   23649    25910: contig of 2262 bp in length
*   25911 26010: gap of      100 bp
*   26011    28060: contig of 2050 bp in length
*   28061 28160: gap of      100 bp
*   28161    30134: contig of 1974 bp in length
*   30135 30234: gap of      100 bp
*   30235    32646: contig of 2412 bp in length
*   32647 32746: gap of      100 bp
*   32747    35096: contig of 2350 bp in length
*   35097 35196: gap of      100 bp
*   35197    37696: contig of 2500 bp in length
*   37697 37796: gap of      100 bp
*   37797    40539: contig of 2743 bp in length
*   40540 40639: gap of      100 bp
*   40640    45518: contig of 4879 bp in length
*   45519 45618: gap of      100 bp
*   45619    48728: contig of 3110 bp in length
*   48729 48828: gap of      100 bp
*   48829    52099: contig of 3271 bp in length
*   52100 52199: gap of      100 bp
*   52200    56747: contig of 4548 bp in length
*   56748 56847: gap of      100 bp
*   56848    60467: contig of 3620 bp in length
*   60468 60567: gap of      100 bp
*   60568    64947: contig of 4380 bp in length
*   64948 65047: gap of      100 bp
*   65048    69431: contig of 4384 bp in length
*   69432 69531: gap of      100 bp
*   69532    74167: contig of 4636 bp in length
*   74168 74267: gap of      100 bp
*   74268    78188: contig of 3921 bp in length
*   78189 78288: gap of      100 bp
*   78289    81664: contig of 3376 bp in length
*   81665 81764: gap of      100 bp
*   81765    85169: contig of 3405 bp in length
*   85170 85269: gap of      100 bp
*   85270    90445: contig of 5176 bp in length
*   90446 90545: gap of      100 bp
*   90546    95008: contig of 4463 bp in length
*   95009 95108: gap of      100 bp
*   95109   100587: contig of 5479 bp in length
*  100588 100687: gap of      100 bp
*  100688   105884: contig of 5197 bp in length
*  105885 105984: gap of      100 bp
*  105985   111486: contig of 5502 bp in length
*  111487 111586: gap of      100 bp
*  111587   117781: contig of 6195 bp in length
*  117782 117881: gap of      100 bp
*  117882   125343: contig of 7462 bp in length
*  125344 125443: gap of      100 bp
*  125444   132508: contig of 7065 bp in length
*  132509 132608: gap of      100 bp
*  132609   141190: contig of 8582 bp in length
*  141191 141290: gap of      100 bp
*  141291   152556: contig of 11266 bp in length
*  152557 152656: gap of      100 bp
*  152657   163590: contig of 10934 bp in length
*  163591 163690: gap of      100 bp
*  163691   176619: contig of 12929 bp in length.
FEATURES             Location/Qualifiers
     source          1. .176619
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11"
                     /clone="RP11-689A10"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1333
                     /note="assembly_fragment"
     misc_feature    1434. .2787
                     /note="assembly_fragment"
     misc_feature    2888. .3998
                     /note="assembly_fragment"
     misc_feature    4099. .5777
                     /note="assembly_fragment"
     misc_feature    5878. .7685
                     /note="assembly_fragment"
     misc_feature    7786. .8176
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:right"
     misc_feature    8277. .9926
                     /note="assembly_fragment"
     misc_feature    10027. .11499
                     /note="assembly_fragment"
     misc_feature    11600. .13616
                     /note="assembly_fragment"
     misc_feature    13717. .16073
                     /note="assembly_fragment"
     misc_feature    16174. .18510
                     /note="assembly_fragment"
     misc_feature    18611. .20495
                     /note="assembly_fragment"
     misc_feature    20596. .23548
                     /note="assembly_fragment"
     misc_feature    23649. .25910
                     /note="assembly_fragment"
     misc_feature    26011. .28060
                     /note="assembly_fragment"
     misc_feature    28161. .30134
                     /note="assembly_fragment"
     misc_feature    30235. .32646
                     /note="assembly_fragment"
     misc_feature    32747. .35096
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
     misc_feature    35197. .37696
                     /note="assembly_fragment"
     misc_feature    37797. .40539
                     /note="assembly_fragment"
     misc_feature    40640. .45518
```

```
                     /note="assembly_fragment"
     misc_feature    45619. .48728
                     /note="assembly_fragment"

  Query Match             100.0%;  Score 20;  DB 70;  Length 176619;
  Best Local Similarity   100.0%;  Pred. No. 0.48;
  Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 gtcacgtgacacagtcaaac 20
          ||||||||||||||||||||
Db  89505 GTCACGTGACACAGTCAAAC 89486


RESULT   8
AF297093
LOCUS       AF297093   198872 bp    DNA             PRI       07-NOV-2000
DEFINITION  Homo sapiens UGT1 gene locus, complete sequence.
ACCESSION   AF297093
VERSION     AF297093.1  GI:11118740
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 198872)
  AUTHORS   Owens,I.S., Gong,Q., Cho,J.W., Huang,T., Potter,C., Gholami,N.,
            Basu,N.K., Kubota,S., Carvalho,S. and Pennington,M.W.
  TITLE     Thirteen UDP glucuronosyltransferase genes encoded at the human
            UGT1 locus
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 198872)
  AUTHORS   Owens,I.S., Gong,Q., Cho,J.W., Huang,T., Potter,C., Gholami,N.,
            Basu,N.K., Kubota,S., Carvalho,S. and Pennington,M.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-AUG-2000) Heritable Disorders Branch, NIH-NICHD, 9000
            Rockville Pike, Bethesda, MD 20892, USA
FEATURES             Location/Qualifiers
     source          1. .198872
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     gene            1989. .2841
                     /gene="UGT1A12p"
                     /note="UDP glucuronosyltransferase 1A12"
                     /pseudo
     CDS             1989. .2841
                     /gene="UGT1A12p"
                     /codon_start=1
                     /pseudo
     gene            19970. .20956
                     /gene="UGT1A11p"
                     /note="UDP glucuronosyltransferase 1A11"
                     /pseudo
     TATA_signal     19970. .19978
     CDS             20105. .20956
                     /gene="UGT1A11p"
                     /codon_start=1
                     /pseudo
     TATA_signal     34133. .34149
                     /gene="UGT1A8"
     gene            34133. .>187313
                     /gene="UGT1A8"
     mRNA            join(<34264. .35118,181788. .181919,182603. .182690,
                     182974. .183193,187016. .>187313)
                     /gene="UGT1A8"
                     /product="UDP glucuronosyltransferase 1A8"
     CDS             join(34264. .35118,181788. .181919,182603. .182690,
                     182974. .183193,187016. .187313)
                     /gene="UGT1A8"
                     /codon_start=1
                     /product="UDP glucuronosyltransferase 1A8"
                     /protein_id="AAG30416.1"
                     /db_xref="GI:11118741"
                     /translation="MARTGWTSPIPLCVSLLLTCGFAEAGKLLVVPMDGSHWFTMQSV
                     VEKLILRGHEVVVVMPEVSWQLGKSLNCTVKTYSTSYTLEDLDREFMDFADAQWKAQV
                     RSLFSLFLSSSNGFFNLFFSHCRSLFNDRKLVEYLKESSFDAVFLDPFDACGLIVAKY
                     FSLPSVVFARGIACHYLEEGAQCPAPLSYVPRILLGFSDAMTFKERVRNHIMHLEEHL
                     FCQYFSKNALEIASEILQTPVTAYDLYSHTSIWLLRTDFVLDYPKPVMPNMIFIGGIN
                     CHQGKPLPMEFEAYINASGEHGIVVFSLGSMVSEIPEKKAMAIADALGKIPQTVLWRY
                     TGTRPSNLANNTILVKWLPQNDLLGHPMTRAFITHAGSHGVYESICNGVPMVMMPLFG
                     DQMDNAKRMETKGAGVTLNVLEMTSEDLENALKAVINDKSYKENIMRLSSLHKDRPVE
                     PLDLAVFWVEFVMRHKGAPHLRPAAHDLTWYQYHSLDVIGFLLAVVLTVAFITFKCCA
                     YGYRKCLGKKGRVKKAHKSKTH"
     TATA_signal     52972. .52988
                     /gene="UGT1A10"
     gene            52972. .>187313
                     /gene="UGT1A10"
     mRNA            join(<53103. .53957,181788. .181919,182603. .182690,
                     182974. .183193,187016. .>187313)
                     /gene="UGT1A10"
                     /product="UDP glucuronosyltransferase 1A10"
     CDS             join(53103. .53957,181788. .181919,182603. .182690,
                     182974. .183193,187016. .187313)
                     /gene="UGT1A10"
                     /codon_start=1
                     /product="UDP glucuronosyltransferase 1A10"
                     /protein_id="AAG30417.1"
                     /db_xref="GI:11118742"
                     /translation="MARAGWTSPVPLCVCLLLTCGFAEAGKLLVVPMDGSHWFTMQSV
                     VEKLILRGHEVVVVMPEVSWQLERSLNCTVKTYSTSYTLEDQNREFMVFAHAQWKAQA
                     QSIFSLLMSSSSGFLDLFFSHCRSLFNDRKLVEYLKESSFDAVFLDPFDTCGLIVAKY
                     FSLPSVVFTRGIFCHHLEEGAQCPAPLSYVPNDLLGFSDAMTFKERVWNHIVHLEDHL
                     FCQYLFRNALEIASEILQTPVTAYDLYSHTSIWLLRTDFVLDYPKPVMPNMIFIGGIN
                     CHQGKPLPMEFEAYINASGEHGIVVFSLGSMVSEIPEKKAMAIADALGKIPQTVLWRY
                     TGTRPSNLANNTILVKWLPQNDLLGHPMTRAFITHAGSHGVYESICNGVPMVMMPLFG
                     DQMDNAKRMETKGAGVTLNVLEMTSEDLENALKAVINDKSYKENIMRLSSLHKDRPVE
                     PLDLAVFWVEFVMRHKGAPHLRPAAHDLTWYQYHSLDVIGFLLAVVLTVAFITFKCCA
                     YGYRKCLGKKGRVKKAHKSKTH"
     gene            64509. .65608
                     /gene="UGT1A13p"
                     /note="UDP glucuronosyltransferase 1A13"
                     /pseudo
     CDS             join(64509. .65011,65353. .65608)
                     /gene="UGT1A13p"
                     /codon_start=1
                     /pseudo
     gene            88416. .>187313
                     /gene="UGT1A9"
     TATA_signal     88416. .88427
                     /gene="UGT1A9"
     mRNA            join(<88543. .89397,181788. .181919,182603. .182690,
                     182974. .183193,187016. .>187313)
                     /gene="UGT1A9"
                     /product="UDP glucuronosyltransferase 1A9"
     CDS             join(88543. .89397,181788. .181919,182603. .182690,
                     182974. .183193,187016. .187313)
                     /gene="UGT1A9"
                     /codon_start=1
                     /product="UDP glucuronosyltransferase 1A9"
                     /protein_id="AAG30418.1"
                     /db_xref="GI:11118743"
                     /translation="MACTGWTSPLPLCVCLLLTCGFAEAGKLLVVPMDGSHWFTMRSV
                     VEKLILRGHEVVVVMPEVSWQLGRSLNCTVKTYSTSYTLEDLDREFKAFAHAQWKAQV
                     RSIYSLLMGSYNDIFDLFFSNCRSLFKDKKLVEYLKESSFDAVFLDPFDNCGLIVAKY
                     FSLPSVVFARGILCHYLEEGAQCPAPLSYVPRILLGFSDAMTFKERVRNHIMHLEEHL
                     LCHRFFKNALEIASEILQTPVTEYDLYSHTSIWLLRTDFVLDYPKPVMPNMIFIGGIN
                     CHQGKPLPMEFEAYINASGEHGIVVFSLGSMVSEIPEKKAMAIADALGKIPQTVLWRY
                     TGTRPSNLANNTILVKWLPQNDLLGHPMTRAFITHAGSHGVYESICNGVPMVMMPLFG
                     DQMDNAKRMETKGAGVTLNVLEMTSEDLENALKAVINDKSYKENIMRLSSLHKDRPVE
                     PLDLAVFWVEFVMRHKGAPHLRPAAHDLTWYQYHSLDVIGFLLAVVLTVAFITFKCCA
                     YGYRKCLGKKGRVKKAHKSKTH"
     TATA_signal     98424. .98438
                     /gene="UGT1A7"
     gene            98424. .>187313
                     /gene="UGT1A7"
     mRNA            join(<98552. .99406,181788. .181919,182603. .182690,
                     182974. .183193,187016. .>187313)
```

```
                     /gene="UGT1A7"
                     /product="UDP glucuronosyltransferase 1A7"
     CDS             join(98552. .99406,181788. .181919,182603. .182690,
                     182974. .183193,187016. .187313)
                     /gene="UGT1A7"
                     /codon_start=1
                     /product="UDP glucuronosyltransferase 1A7"
                     /protein_id="AAG30419.1"
                     /db_xref="GI:11118744"
                     /translation="MARAGWTGLLPLYVCLLLTCGFAKAGKLLVVPMDGSHWFTMQSV
                     VEKLILRGHEVVVVMPEVSWQLGRSLNCTVKTYSTSYTLEDQDREFMVFADARWTAPL
                     RSAFSLLTSSSNGIFDLFFSNCRSLFKDKKLVEYLKESCFDAVFLDPFDACGLIVAKY
                     FSLPSVVFARGIFCHYLEEGAQCPAPLSYVPRLLLGFSDAMTFKERVRNHIMHLEEHL
                     FCPYFFKNVLEIASEILQTPVTAYDLYSHTSIWLLRTDFVLEYPKPVMPNMIFIGGIN
                     CHQGKPVPMEFEAYINASGEHGIVVFSLGSMVSEIPEKKAMAIADALGKIPQTVLWRY
                     TGTRPSNLANNTILVKWLPQNDLLGHPMTRAFITHAGSHGVYESICNGVPMVMMPLFG
                     DQMDNAKRMETKGAGVTLNVLEMTSEDLENALKAVINDKSYKENIMRLSSLHKDRPVE
                     PLDLAVFWVEFVMRHKGAPHLRPAAHDLTWYQYHSLDVIGFLLAVVLTVAFITFKCCA
                     YGYRKCLGKKGRVKKAHKSKTH"
     gene            109467. .>187313
                     /gene="UGT1A6"
     TATA_signal     109467. .109473
                     /gene="UGT1A6"
     mRNA            join(<109610. .110470,181788. .181919,182603. .182690,
                     182974. .183193,187016. .>187313)
                     /gene="UGT1A6"
                     /product="UDP glucuronosyltransferase 1A6"
     CDS             join(109610. .110470,181788. .181919,182603. .182690,
                     182974. .183193,187016. .187313)
                     /gene="UGT1A6"
                     /codon_start=1
                     /product="UDP glucuronosyltransferase 1A6"
                     /protein_id="AAG30420.1"
                     /db_xref="GI:11118745"
                     /translation="MACLLRSFQRISAGVFFLALWGMVVGDKLLVVPQDGSHWLSMKD
                     IVEVLSDRGHEIVVVVPEVNLLLKESKYYTRKIYPVPYDQEELKNRYQSFGNNHFAER
                     SFLTAPQTEYRNNMIVIGLYFINCQSLLQDRDTLNFFKESKFDALFTDPALPCGVILA
                     EYLGLPSVYLFRGFPCSLEHTFSRSPDPVSYIPRCYTKFSDHMTFSQRVANFLVNLLE
                     PYLFYCLFSKYEELASAVLKRDVDIITLYQKVSVWLLRYDFVLEYPRPVMPNMVFIGG
                     INCKKRKDLSQEFEAYINASGEHGIVVFSLGSMVSEIPEKKAMAIADALGKIPQTVLW
                     RYTGTRPSNLANNTILVKWLPQNDLLGHPMTRAFITHAGSHGVYESICNGVPMVMMPL
                     FGDQMDNAKRMETKGAGVTLNVLEMTSEDLENALKAVINDKSYKENIMRLSSLHKDRP
                     VEPLDLAVFWVEFVMRHKGAPHLRPAAHDLTWYQYHSLDVIGFLLAVVLTVAFITFKC
                     CAYGYRKCLGKKGRVKKAHKSKTH"
     gene            129525. .>187313
                     /gene="UGT1A5"
     TATA_signal     129525. .129531
                     /gene="UGT1A5"
     mRNA            join(<129600. .130466,181788. .181919,182603. .182690,
                     182974. .183193,187016. .>187313)
                     /gene="UGT1A5"
                     /product="UDP glucuronosyltransferase 1A5"
     CDS             join(129600. .130466,181788. .181919,182603. .182690,
                     182974. .183193,187016. .187313)
                     /gene="UGT1A5"
                     /codon_start=1
                     /product="UDP glucuronosyltransferase 1A5"
                     /protein_id="AAG30421.1"
                     /db_xref="GI:11118746"
                     /translation="MATGLQVPLPQLATGLLLLLSVQPWAESGKVLVVPTDGSHWLSM
                     REALRDLHARGHQVVVLTLEVNMYIKEENFFTLTTYAISWTQDEFDRLLLGHTQSFFE
                     TEHLLMKFSRRMAIMNNMSLIIHRSCVELLHNEALIRHLHATSFDVVLTDPFHLCAAV
                     LAKYLSIPAVFFLRNIPCDLDFKGTQCPNPSSYIPRLLTTNSDHMTFLQRVKNMLYPL
                     ALSYLCHAVSAPYASLASELFQREVSVVDLVSHASVWLFRGDFVMDYPRPIMPNMVFI
                     GGINCANGKPLSQEFEAYINASGEHGIVVFSLGSMVSEIPEKKAMAIADALGKIPQTV
                     LWRYTGTRPSNLANNTILVKWLPQNDLLGHPMTRAFITHAGSHGVYESICNGVPMVMM
                     PLFGDQMDNAKRMETKGAGVTLNVLEMTSEDLENALKAVINDKSYKENIMRLSSLHKD
                     RPVEPLDLAVFWVEFVMRHKGAPHLRPAAHDLTWYQYHSLDVIGFLLAVVLTVAFITF
                     KCCAYGYRKCLGKKGRVKKAHKSKTH"

Query Match           100.0%;  Score 20;  DB 89;  Length 198872;
Best Local Similarity  100.0%;  Pred. No. 0.48;
Matches   20;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 gtcacgtgacacagtcaaac 20
           ||||||||||||||||||||
Db  174939 GTCACGTGACACAGTCAAAC 174958


RESULT   9
AF135466
LOCUS       AF135466      200 bp    DNA             PRI       21-NOV-1999
DEFINITION  Pongo pygmaeus UDP-glucuronosyltransferase 1A1 (UGT1A1) gene,
            promoter region and partial cds.
ACCESSION   AF135466
VERSION     AF135466.1  GI:6456549
KEYWORDS    .
SOURCE      orangutan.
  ORGANISM  Pongo pygmaeus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pongo.
REFERENCE   1  (bases 1 to 200)
  AUTHORS   Hall,D., Ybazeta,G., Destro-Bisol,G., Petzl-Erler,M.L. and Di
            Rienzo,A.
  TITLE     Variability at the uridine diphosphate glucuronosyltransferase 1A1
            promoter in human populations and primates
  JOURNAL   Pharmacogenetics (1999) In press
REFERENCE   2  (bases 1 to 200)
  AUTHORS   Ybazeta,G., Hall,D. and Di Rienzo,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-MAR-1999) Human Genetics, University of Chicago, 924
            E. 57th Street, Chicago, IL 60637, USA
FEATURES             Location/Qualifiers
     source          1. .200
                     /organism="Pongo pygmaeus"
                     /db_xref="taxon:9600"
     mRNA            <121. .>200
                     /gene="UGT1A1"
                     /product="UDP-glucuronosyltransferase 1A1"
     gene            <121. .>200
                     /gene="UGT1A1"
     CDS             121. .>200
                     /gene="UGT1A1"
                     /codon_start=1
                     /product="UDP-glucuronosyltransferase 1A1"
                     /protein_id="AAF09177.1"
                     /db_xref="GI:6456550"
                     /translation="MAVESQGGRPLVLGLLLCVLGPVVSHA"
BASE COUNT       40 a     46 c     62 g     52 t
ORIGIN


  Query Match           92.0%;  Score 18.4;  DB 89;  Length 200;
  Best Local Similarity  95.0%;  Pred. No. 5.6;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 gtcacgtgacacagtcaaac 20
           |||| ||||||||||||||||
Db      23 GTCATGTGACACAGTCAAAC 42


RESULT  10
AF135467
LOCUS       AF135467      208 bp    DNA             PRI       21-NOV-1999
DEFINITION  Trachypithecus obscurus UDP-glucuronosyltransferase 1A1 (UGT1A1)
            gene, promoter region and partial cds.
ACCESSION   AF135467
VERSION     AF135467.1  GI:6456551
KEYWORDS    .
SOURCE      dusky leaf monkey.
  ORGANISM  Trachypithecus obscurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
            Colobinae; Trachypithecus.
REFERENCE   1  (bases 1 to 208)
  AUTHORS   Hall,D., Ybazeta,G., Destro-Bisol,G., Petzl-Erler,M.L. and Di
```

```
            Rienzo,A.
  TITLE     Variability at the uridine diphosphate glucuronosyltransferase 1A1
            promoter in human populations and primates
  JOURNAL   Pharmacogenetics (1999) In press
REFERENCE   2  (bases 1 to 208)
  AUTHORS   Ybazeta,G., Hall,D. and Di Rienzo,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-MAR-1999) Human Genetics, University of Chicago, 924
            E. 57th Street, Chicago, IL 60637, USA
FEATURES             Location/Qualifiers
     source          1. .208
                     /organism="Trachypithecus obscurus"
                     /db_xref="taxon:54181"
     mRNA            <125. .>208
                     /gene="UGT1A1"
                     /product="UDP-glucuronosyltransferase 1A1"
     gene            <125. .>208
                     /gene="UGT1A1"
     CDS             125. .>208
                     /gene="UGT1A1"
                     /codon_start=1
                     /product="UDP-glucuronosyltransferase 1A1"
                     /protein_id="AAF09178.1"
                     /db_xref="GI:6456552"
                     /translation="MAVESQDRRPLVLGLLLCVLGPVVCHAG"
BASE COUNT       45 a     46 c     60 g     57 t
ORIGIN


  Query Match           92.0%;  Score 18.4;  DB 89;  Length 208;
  Best Local Similarity  95.0%;  Pred. No. 5.6;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 gtcacgtgacacagtcaaac 20
          |||||||||||||| ||||||
Db     30 GTCACGTGACACAATCAAAC 49


RESULT  11
AF135462
LOCUS       AF135462      234 bp    DNA             PRI       21-NOV-1999
DEFINITION  Pan paniscus UDP-glucuronosyltransferase 1A1 (UGT1A1) gene,
            promoter region and partial cds.
ACCESSION   AF135462
VERSION     AF135462.1  GI:6456541
KEYWORDS    .
SOURCE      pygmy chimpanzee.
  ORGANISM  Pan paniscus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1  (bases 1 to 234)
  AUTHORS   Hall,D., Ybazeta,G., Destro-Bisol,G., Petzl-Erler,M.L. and Di
            Rienzo,A.
  TITLE     Variability at the uridine diphosphate glucuronosyltransferase 1A1
            promoter in human populations and primates
  JOURNAL   Pharmacogenetics (1999) In press
REFERENCE   2  (bases 1 to 234)
  AUTHORS   Ybazeta,G., Hall,D. and Di Rienzo,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-MAR-1999) Human Genetics, University of Chicago, 924
            E. 57th Street, Chicago, IL 60637, USA
FEATURES             Location/Qualifiers
     source          1. .234
                     /organism="Pan paniscus"
                     /db_xref="taxon:9597"
     mRNA            <148. .>234
                     /gene="UGT1A1"
                     /product="UDP-glucuronosyltransferase 1A1"
     gene            <148. .>234
                     /gene="UGT1A1"
     CDS             148. .>234
                     /gene="UGT1A1"
                     /codon_start=1
                     /product="UDP-glucuronosyltransferase 1A1"
                     /protein_id="AAF09173.1"
                     /db_xref="GI:6456542"
                     /translation="MAVESQGGRPLVLGLLLCVLGPVVSHAGK"
BASE COUNT       50 a     53 c     69 g     62 t
ORIGIN


  Query Match           92.0%;  Score 18.4;  DB 89;  Length 234;
  Best Local Similarity  95.0%;  Pred. No. 5.6;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 gtcacgtgacacagtcaaac 20
          |||| ||||||||||||||||
Db     46 GTCATGTGACACAGTCAAAC 65


RESULT  12
AF135463
LOCUS       AF135463      234 bp    DNA             PRI       21-NOV-1999
DEFINITION  Pan troglodytes UDP-glucuronosyltransferase 1A1 (UGT1A1) gene,
            promoter region and partial cds.
ACCESSION   AF135463
VERSION     AF135463.1  GI:6456543
KEYWORDS    .
SOURCE      chimpanzee.
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1  (bases 1 to 234)
  AUTHORS   Hall,D., Ybazeta,G., Destro-Bisol,G., Petzl-Erler,M.L. and Di
            Rienzo,A.
  TITLE     Variability at the uridine diphosphate glucuronosyltransferase 1A1
            promoter in human populations and primates
  JOURNAL   Pharmacogenetics (1999) In press
REFERENCE   2  (bases 1 to 234)
  AUTHORS   Ybazeta,G., Hall,D. and Di Rienzo,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-MAR-1999) Human Genetics, University of Chicago, 924
            E. 57th Street, Chicago, IL 60637, USA
FEATURES             Location/Qualifiers
     source          1. .234
                     /organism="Pan troglodytes"
                     /db_xref="taxon:9598"
     mRNA            <148. .>234
                     /gene="UGT1A1"
                     /product="UDP-glucuronosyltransferase 1A1"
     gene            <148. .>234
                     /gene="UGT1A1"
     CDS             148. .>234
                     /gene="UGT1A1"
                     /codon_start=1
                     /product="UDP-glucuronosyltransferase 1A1"
                     /protein_id="AAF09174.1"
                     /db_xref="GI:6456544"
                     /translation="MAVESQGGRPLVLGLLLCVLGPVVSHAGK"
BASE COUNT       50 a     53 c     69 g     62 t
ORIGIN


  Query Match           92.0%;  Score 18.4;  DB 89;  Length 234;
  Best Local Similarity  95.0%;  Pred. No. 5.6;
  Matches   19;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 gtcacgtgacacagtcaaac 20
          |||| ||||||||||||||||
Db     46 GTCATGTGACACAGTCAAAC 65


RESULT  13
MUSPC326A
LOCUS       MUSPC326A     4201 bp    mRNA            ROD       27-APR-1993
DEFINITION  Mus musculus protein PC326 mRNA, complete cds.
```

```
ACCESSION   M95564
VERSION     M95564.1  GI:200240
KEYWORDS    protein PC326.
SOURCE      Mus musculus (strain BALB/c, sub_species domesticus) plasmacytoma
            cDNA to mRNA.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 4201)
  AUTHORS   Eckhardt,L.A., Laskov,R., Kuehl,W.M., Bergsagel,P.L. and
            Timblin,C.R.
  TITLE     Sequence and expression of a murine cDNA encoding PC326, a novel
            gene expressed in plasmacytomas but not normal plasma cells
  JOURNAL   Oncogene 7, 2059-2064 (1992)
  MEDLINE   93026383
FEATURES             Location/Qualifiers
     source          1. .4201
                     /organism="Mus musculus"
                     /strain="BALB/c"
                     /sub_species="domesticus"
                     /db_xref="taxon:10090"
                     /cell_line="4N2.1"
                     /cell_type="plasma cell"
                     /tissue_type="plasmacytoma"
     repeat_region   <1. .1400
                     /rpt_family="B1 repeat"
     gene            1514. .4201
                     /gene="protein PC326"
     CDS             1514. .3757
                     /gene="protein PC326"
                     /note="'member of beta transducin repeat 'mosaic' protein
                     family'; putative"
                     /codon_start=1
                     /product="protein PC326"
                     /protein_id="AAA39895.1"
                     /db_xref="GI:200241"
                     /translation="MSSHESYTNAAETPENISILSCLGETSGALVDTKTISDIKTMDP
                     RVSLTPSSDVTGTEDSSVLTPQSTDVNSVDSYQGYEGDDDDEEDDEDDKDGDSNLPSL
                     EDSDNFISCLENSYIPQNVENGEVVEEQSLGRRFHPYELEAGEVVEGQGGGSLFYPYE
                     LEAGEVVEAQNVQNLFHRYELEEGEVVEAQVVQSMFPYYELEAGEVVEAEEVQGFFQR
                     YELEAREVIGAQGGQGLSRHYGLEGGEVVEATAVRRLIQHHELEEGEDVDDQEESSEM
                     HEETSEDSSEQYDIEDDSLIDEWIALETSPLPRPRWNVLSALRDRQLGSSGRFVYEAC
                     GARLFVQRFSLEHVFEGHSGCVNTVHFNQHGTLLASGSDDLKVIVWDWLKKRSVLNFD
                     SGHKNNILQAKFLPNCNDAILAMCGRDGQVRVAQLSAVAGTHMTKRLVKHGGASHRLG
                     LEPDSPFRFLTSGEDAVVFNIDLRQAHPASKLLVIKDGDKKVGLYTVFVNPANVYQFA
                     VGGQDQFMRIYDQRKIDENVNNGVLKKFCPHHLLSSDYPAHITSLMYSYDGTEILASY
                     NDEDIYIFNSSDSDGAQYAKRYKGHRNNSTVKGVYFYGPRSEFVMSGSDCGHIFIWEK
                     SSCQIVQFLEADEGGTINCIDSHPYLPVLASSGLDHEVKIWSPIAEPSKKLAGLKNVI
                     KINKLKRDNFTLRHTSLFNNSMLCFLMSHVTQSNYGRSWRGIRINAGGGDFSDSSSSS
                     EETNQES"
     repeat_region   1882. .2304
                     /standard_name="'PC326'"
                     /note="'20 amino acid repeat'; putative"
                     /rpt_type=tandem
     3'UTR           3575. .4201
                     /gene="protein PC326"
                     /note="putative"
     polyA_signal    4168. .4174
                     /gene="protein PC326"
                     /note="putative"
     polyA_site      4186
                     /gene="protein PC326"
                     /note="putative"
BASE COUNT     1331 a    755 c   1006 g   1101 t      8 others
ORIGIN


  Query Match           87.0%;  Score 17.4;  DB 94;  Length 4201;
  Best Local Similarity  94.7%;  Pred. No. 20;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 gtcacgtgacacagtcaaa 19
          |||| ||||||||||||||
```

```
Db   3642 GTCATGTGACACAGTCAAA 3660


RESULT 14
AF069716/c
LOCUS       AF069716     99123 bp    DNA             PLN       11-JUN-1998
DEFINITION  Arabidopsis Thaliana BAC F6A4, Chromosome IV, near 60.5 cM,
            complete sequence.
ACCESSION   AF069716
VERSION     AF069716.1  GI:3212102
KEYWORDS    HTG.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
            Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 99123)
  AUTHORS   Parnell,L.D., Huang,E.N., de la Bastide,M., Calma,C., Habermann,K.,
            Schutz,K., Gnoj,L., Dedhia,N.N. and Simon,N.
  TITLE     Genomic Sequence of Arabidopsis Thaliana BAC F6A4, Chromosome IV,
            near 60.5 cM
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 99123)
  AUTHORS   Parnell,L.D., Huang,E.N., de la Bastide,M., Calma,C., Habermann,K.,
            Schutz,K., Gnoj,L., Dedhia,N.N. and Simon,N.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-JUN-1998) Lita Annenberg Hazen Genome Center, Cold
            Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, NY
            11724, USA
COMMENT     BAC F6A4 was sequenced as part of the arabidopsis genome sequencing
            effort of the Cold Spring Harbor Consortium. For additional
            information, please see http://www.cshl.org/arabweb. F6A4 overlaps
            with T11H3 towards the centromere. Fingerprint data place F6A4
            about 150 kbp to the left of the KNAT3 marker.
FEATURES             Location/Qualifiers
     source          1. .99123
                     /organism="Arabidopsis thaliana"
                     /cultivar="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="V"
                     /map="60.5 cM"
                     /clone="F6A4"
                     /note="hybridizing YAC is CIC2E3"
BASE COUNT     31735 a   16898 c   17599 g   32891 t
ORIGIN


  Query Match           87.0%;  Score 17.4;  DB 12;  Length 99123;
  Best Local Similarity  94.7%;  Pred. No. 18;
  Matches   18;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      2 tcacgtgacacagtcaaac 20
          |||||| ||||||||||||
Db  49925 TCACGTAACACAGTCAAAC 49907


RESULT 15
AF135470
LOCUS       AF135470      207 bp    DNA             PRI       21-NOV-1999
DEFINITION  Cebus apella UDP-glucuronosyltransferase 1A1 (UGT1A1) gene,
            promoter region and partial cds.
ACCESSION   AF135470
VERSION     AF135470.1  GI:6456557
KEYWORDS    .
SOURCE      brown capuchin.
  ORGANISM  Cebus apella
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Platyrrhini; Cebidae; Cebinae; Cebus.
REFERENCE   1  (bases 1 to 207)
  AUTHORS   Hall,D., Ybazeta,G., Destro-Bisol,G., Petzl-Erler,M.L. and Di
            Rienzo,A.
  TITLE     Variability at the uridine diphosphate glucuronosyltransferase 1A1
            promoter in human populations and primates
```

```
  JOURNAL   Pharmacogenetics (1999) In press
REFERENCE   2  (bases 1 to 207)
  AUTHORS   Ybazeta,G., Hall,D. and Di Rienzo,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-MAR-1999) Human Genetics, University of Chicago, 924
            E. 57th Street, Chicago, IL 60637, USA
FEATURES             Location/Qualifiers
     source          1. .207
                     /organism="Cebus apella"
                     /db_xref="taxon:9515"
     mRNA            <144. .>207
                     /gene="UGT1A1"
                     /product="UDP-glucuronosyltransferase 1A1"
     gene            <144. .>207
                     /gene="UGT1A1"
     CDS             144. .>207
                     /gene="UGT1A1"
                     /codon_start=1
                     /product="UDP-glucuronosyltransferase 1A1"
                     /protein_id="AAF09181.1"
                     /db_xref="GI:6456558"
                     /translation="MPACPGPAAVCAGPGSVPCWE"
BASE COUNT       43 a     46 c     62 g     56 t
ORIGIN


  Query Match             84.0%;  Score 16.8;  DB 89;  Length 207;
  Best Local Similarity   90.0%;  Pred. No. 52;
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 gtcacgtgacacagtcaaac 20
           |||| |||||||| ||||||
Db      23 GTCA[illegible]TGACACAATCAAAC 42


Search completed: July 25, 2001, 05:17:09
Job time: 9232 sec
```